

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:58:11 ; Search time 90.5 Seconds
(without alignments)
12118.688 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179
Sequence: 1 99cggaggcggagcggagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p_model -DEV=xlp
-Q=/cgn2/1/USPRO_spool/p/US09323597/runat_06072004_111728_29658/app_query.fasta_1.1927
-DB=SPREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597 @CNG 1 1 99 @runat_06072004_111728_29658 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DB_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL 25:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_nhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp Vertebrate:.*
14: sp Unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2711	85.3	492 4	Q96T73 homo sapien

RESULT 1

ID	Q96T73	PRELIMINARY:	PRT;	492 AA.
AC	Q96T73;			
DT	01-DEC-2001	{T-EMBLrel. 19, Created}		
DT	01-DEC-2001	{T-EMBLrel. 19, Last sequence update}		
DT	01-OCT-2003	{T-EMBLrel. 25, Last annotation update}		
DE	Epitheliasin.			
GN	TMPRSS2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21223025; PubMed=11322890;			
RA	Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,			
RA	Hoidal J.R.;			
RT	"Cloning and characterization of the cDNA and gene for human			
RT	epitheliasin."			
EL	Eur. J. Biochem. 268:2687-2699(2001).			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	EMBL; AF329454; AAK5359.1; -			
DR	HSSP; P09761; IAN1.			
DR	GO; GO:0016020; C:membrane; IEA.			

ALIGNMENTS

	2	2121	66.7	490	11	Q7TN04
3	2083	65.5	490	11	Q920K3	Q7tn04 mus musculus
4	1009	31.7	767	13	Q9DGR2	Q920k3 rattus norv
5	881	27.7	453	11	Q812A6	Q9dgr2 xenopus lae
6	751	23.6	537	4	Q9BYE1	Q812a6 mus musculus
7	736.5	23.2	591	4	Q9BYE2	Q9bye1 homo sapien
8	731	23.0	471	11	Q8CFE0	Q9bye2 homo sapien
9	725.5	22.8	558	4	Q86YM4	Q8cfe0 mus musculus
10	676.5	21.3	405	4	Q96E86	Q86ym4 homo sapien
11	676.5	21.3	445	11	Q8CJ17	Q96e86 homo sapien
12	664.5	20.9	455	11	Q8CDR0	Q8cj17 rattus norv
13	662.5	20.8	371	11	Q8CJ16	Q8cdro mus musculus
14	649	20.4	226	13	Q7Z280	Q8cj16 rattus norv
15	646.5	20.3	777	11	Q8CAN9	Q7z280 brachydanio
16	602	18.9	1111	11	Q80YN4	Q8can9 mus musculus
17	579	18.2	624	6	Q95ME7	Q80yn4 rattus norv
18	570.5	17.9	643	6	Q97506	Q95me7 oryctolagus
19	564.5	17.8	855	11	Q9J17	Q97506 sus scrofa
20	560.5	17.6	624	11	Q9DAT3	Q9j17 rattus norv
21	558.5	17.6	422	4	Q8WVC1	Q9dat3 mus musculus
22	553	17.4	624	11	Q91Y47	Q8wvc1 homo sapien
23	544.5	17.1	845	13	Q9DGR1	Q91y47 mus musculus
24	544	17.1	1379	5	Q9V4N6	Q9dgr1 xenopus lae
25	542	17.0	638	11	Q8ROP5	Q9v4n6 drosophila
26	541	17.0	572	11	Q8BIK6	Q8rop5 mus musculus
27	531	16.7	310	11	Q9QY29	Q8bik6 mus musculus
28	528	16.6	812	11	Q9R0W3	Q9qy29 mus musculus
29	525	16.5	310	11	Q91XC4	Q9r0w3 rattus norv
30	519.5	16.3	439	11	Q8BHM9	Q91xc4 mus musculus
31	517.5	16.3	329	6	Q9GL10	Q8bhm9 mus musculus
32	516	16.2	417	11	Q8VHJ4	Q9gl10 ovis aries
33	514	16.2	417	11	Q8VDV1	Q8vhj4 rattus norv
34	514	16.2	417	11	Q8VHK8	Q8vdv1 mus musculus
35	511.5	16.1	855	4	Q7Z410	Q8vhk8 mus musculus
36	511.5	16.1	1059	4	Q7Z411	Q7z410 homo sapien
37	509.5	16.0	331	11	Q8RIA6	Q7z411 homo sapien
38	509.5	16.0	389	13	Q9PVX7	Q8ria6 mus musculus
39	509	16.0	417	11	Q8BZ10	Q9pvx7 xenopus lae
40	508.5	16.0	279	11	Q7TNX3	Q8bz10 mus musculus
41	508.5	16.0	416	4	Q86T26	Q7tnx3 mus musculus
42	507.5	16.0	220	13	Q7TOX2	Q86t26 homo sapien
43	506.5	15.9	331	11	Q80X17	Q7tox2 xenopus lae
44	506	15.9	336	11	Q80YD8	Q80x17 mus musculus
45	505	15.9	328	11	Q8BJR6	Q80yd8 mus musculus

DR GO: 0004263; F:chymotrypsin activity; IEA.
DR GO: 0008233; F:peptidase activity; IEA.
DR GO: 0005044; F:scavenger receptor activity; IEA.
DR GO: 0004295; F:trypsin activity; IEA.
DR GO: 0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001344; Peptidase_S1A.
DR InterPro: IPR001190; Sctri_receptor.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLA; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00200; Tryp_SPC; 1.
DR PROSITE: PS01209; LDLA_1; 1.
DR PROSITE: PS00686; LDLA_2; 1.
DR PROSITE: PS0287; SRCR_2; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Alignment Scores:
Pred. No.: 2,76e-230 Length: 492
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 85.28% Indels: 0
DB: 4 Gaps: 0

US-09-323-597c-1 (1-1738) x Q96T73 (1-492)
QY 112 ATGGCTTTGAACCTCAGGTCACACAGCTATTGGACCTTACTATGAAACACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAAACCGAAACCCCTATCCGACAGCCCACTGTGTCCTCCCACTGTCTACAGGTCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGGCTCAGTACTACCGCTCCCGCGGCCAGTACGCCCGAGGCTCTGACGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCGCTCTGTCACAGCAGCCCAATCCCATCCGGGACAGTGTGACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAGAACCACTGTGCATCACCCTGACCTGGGACCTTCTCTGGGAGCTCGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAAGTTTCATGGGACAGTGTCTCAACTCTGGGATAGATGC 471
DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACTGTCATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAACTCGTGTGTTCCCTCTACGGACCAACTTCACTCTCAGGTG 591
DB 141 GlyGlyLysGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGGAAGCTCTGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GGGCGGGCGCTGACGGGACATGGCTATAGATAATTTTACTCTAGCCCAAGGATA 711
DB 181 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPherySerSerGlnGlyIle 200

QY 712 GTGGATGACAGGGATCCACAGCTTTATGAACTGAACACAGTGCCTGGCAATTCGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTACACAGTGTATCCCTGCTTCTCAAAGACAGTGTCTTCTTTCAGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCCCGCAGAGAGAGATTGTGGCGCGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCTGGCCCTGGCAGGTTCAGCTGCAGCTGCAGAACTGCACAGT 951
DB 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCATCCCCCGAGTGGATCGTGACAGCGGCCCATCTGCTGGA 1011
DB 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGATGGAGCGCANTTGGGGGATTTTGACATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGGGATACCAAGTAGAAAAAGTGTCTTCTCATCCAAATTTATGACTCC 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGACATGACATTCGCTGATGAGCTGCGAAGCTCTGACTTTCAACGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACAGTGTGTCTGCCCAACCCAGGCATGTGTGTCAGCCAGCCAGAACAGCTCT 1251
DB 361 ValLysProValCysLeuLeuProAsnProGlyMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGAGAGAAAGGAGACCTCAGAAAGTCTGAACCTGCC 1311
DB 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATGTAGACACAGATGCAACAGCAGATGTCTATGACAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCCATGTCTGTGGCTCTCTGCGGGAACCTGTCAGGGAACCTGCTTTCAGG 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGGTCACTTCGAAGAACATATCTGGTGGTCTGATAGGGGATACAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGCTGTGCCAAAGCTTACAGACAGAGTGTACGGGAATGTGATGTATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTATCGCAAAATGAGGCGCAGACGCC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2
Q7TN04
ID Q7TN04 PRELIMINARY; PRT; 490 AA.
AC Q7TN04; Q7TN04; PRELIMINARY; PRT; 490 AA.
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trpms2 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RN SEQUENCE FROM N.A.
 RP STRAIN=129; TISSUE=Breast tumor;
 RC MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan K.C., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udino T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Eulky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.B., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.B.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RL prsc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=129; TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC054348; AAH54348.1; -- 546508028417665A CRC64;
 SQ SEQUENCE 490 AA; 53525 MW; 546508028417665A CRC64;

Alignment Scores:

Pred. No.: 3.2e-178 Length: 490
 Score: 2121.00 Matches: 385
 Percent Similarity: 87.17% Conservative: 43
 Best local Similarity: 78.41% Mismatches: 61
 Query Match: 66.72% Indels: 2
 DB: 11 Gaps: 2

US-09-323-597C-1 (1-1738) x Q7TN04 (1-490)

QY 112 ATGGCTTTGAATCACTAGGCTGACCACTGATTTGGACCTTACTATGAAACCATGGATAC 171
 DB 1 MetAlaLeuAsnSerGlySerProGlyIleGlyProCysTyrGluAsnHisGlyTyr 20
 QY 172 CAACGGGAAACCCCTATCCGACACAGCCACTGTGTCCTCCACTGTCTACAGGTGCAT 231
 DB 21 GlnSerGluHisIleCysProProArgProValAlaProAsnGlyTyrAsnLeuTyr 40
 QY 232 CCGCTCAGTACTACCTGCTGCCCGTCCCGTCCAGTACGCGGGTCTCTGACCGAGCT 291
 DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgIleThrThrGlnAla 60
 QY 292 TCCAAACCCCGTCTGTCAGCAGCAGCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 351
 DB 61 SerThrSerValIleHisThrHisProLysSer--SerGlyAlaLeuCysThrSerLys 79
 QY 352 ACTAAGAAGCACTGTGCATCACTTACCTTGGGACCTTCTCTGGGAGCTGGCTG 411
 DB 80 SerLysLysSerLeuCysLeuAlaLeuAlaLeuGlyThrValThrGlyAlaAlaVal 99
 QY 412 GCCGTGGCTTACTCTGGAAGTTCTATGGGCGAGCAAGTCTCCAACTCTGGGATAGATGC 471
 DB 100 AlaAlaValLeuLeuTyrArgPheThrAspSerAsnCysSerThrSerGluMetGluCys 119
 QY 472 GACTCTCAGGTACCTGCATCAACCCCTTCACTTACCTGTGTGTGTGTGTGTGTGTGTGT 531
 DB 120 GlySerSerGlyThrCysIleSerSerSerSerSerSerSerSerSerSerSerSerSer 139
 QY 532 GCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 591
 DB 140 AsnGlyGluAspGluAsnArgCysValArgLeuTyrGlyGlnSerPheIleLeuGlnVal 159

RESULT 3

Q920K3
 ID Q920K3
 AC Q920K3;

PRELIMINARY; PRT; 490 AA.

QY 592 TACTCATCTCAGAGGAGTCTCTGGCACCCTGTGTGCCAAGCACTGGAACGAGAACTAC 651
 DB 160 TyrSerSerGlnArgLysAlaIleTyrProValCysGlnAspSerTrpSerGluSerTyr 179
 QY 652 GGGCGGGCGGCTGCGCAGGAGCATGGGTATATAAGATAATTTTACTCTAGCCAGGAATA 711
 DB 180 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 199
 QY 712 GTGGATGACAGCGGATCCACACCTTTATGAAATGAAACCAAGTCCGCGCAATGTGAT 771
 DB 200 ProAspGlnSerGlyAlaThrSerPheMetLysLeuAsnValSerSerGlyAsnValAsp 219
 QY 772 ATCTATAAAAACTGTACACACAGTGTGCTGTTCTTCAAAAGCAGTGGTTCCTTACGC 831
 DB 220 LeuTyrLysLysLeuTyrHisSerAspSerCysSerSerArgMetValValSerLeuArg 239
 QY 832 TGTATACCTTGGCGGGTCAACTTGAACCTAAGCCGCCAGAGCGAGGATTTGGCGGCGGAG 891
 DB 240 CysIleGluCysGlyValArg--SerValLysArgGlnSerArgIleValGlyGlyLeu 258
 QY 892 AGCGCGTCCCGGGGGCTGGCCCTGGCAGGTGACCTGCACCTGCCAAGACGTCCACGTG 951
 DB 259 AsnAlaSerProGlyAspTrpProTrpGlnValSerLeuHisValGlnGlyValHisVal 278
 QY 952 TGGGAGGCTCCATCATCACCCCGAGTGTGATCGTGACAGCCGCCACTCGGTGAAAAA 1011
 DB 279 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluGlu 298
 QY 1012 CCTCTTAACAATCCATGGCATTTGAGCGGCAATTTCCGGGGATTTTGACACATCTTTTCATG 1071
 DB 299 ProLeuSerSerProArgTyrTrpThrAlaPheAlaGlyIleLeuArgGlnSerLeuMet 318
 QY 1072 TTCTATGGACCGGATACCAAGTAGAAAAAGTGTATTTCTCATCCAAATTATGATCCCAAG 1131
 DB 319 PheTyrGlySerArgHisGlnValGluLysValIleSerHisProAsnTyrAspSerLys 338
 QY 1132 ACAGAGCAATGACATTTGGCTGTGATGACAGCTGCAGAGGCTCTGACTTTCAACGACCTA 1191
 DB 339 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnThrProLeuAlaPheAsnAspLeu 358
 QY 1192 GTGAAACCACTGTGTCTGCCCAACCCAGGATGATGTGTCAGCCAGCAACAGCTCTGTGG 1251
 DB 359 ValLysProValCysLeuProAsnProGlyMetMetLeuAspLeuAspGlnGluCysTrp 378
 QY 1252 ATTTCCGGTGGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAAAGTGTGTGACGCTGCC 1311
 DB 379 IleSerGlyTrpGlyAlaThrTyrGluLysGlyLysThrSerAspValLeuAsnAlaAla 398
 QY 1312 AAGGTCTTCTCATTGAGACACAGATGCAACAGCAGATATGCTATGACCAACTGATC 1371
 DB 399 MetValProLeuIleGluProSerLysCysAsnSerLysTyrIleTyrAsnAsnLeuIle 418
 QY 1372 ACACAGCCCATGATCTGTGCGGCTTCTGCGAGGAAACCTGCGATTTCTGCGAGGTGAC 1431
 DB 419 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlySerValAspSerCysGlnGlyAsp 438
 QY 1432 AGTGGAGGCTCTGGTCACTTCGAGAACAATATCTGGTGGCTGTAGGGGATACAGC 1491
 DB 439 SerGlyGlyProLeuValThrLeuLysAsnGlyIleTrpIleLeuIleGlyAspThrSer 458
 QY 1492 TGGGGTTCTGGCTGTGCAAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTC 1551
 DB 459 TrpGlySerGlyCysAlaLysAlaLeuArgProGlyValTyrGlyAsnValThrValPhe 478
 QY 1552 ACGACATGGATTTATCGACAAATGAGGGCAGAC 1584
 DB 479 ThrAspTrpIleTyrGlnGlnMetArgAlaAsn 489

Qy	1552	ACGACCTGGATTATTCGACAAATAGGGGAGAC	1584
Db	479	ThrAspTrpIleTyrGlnGlnMetArgAlaAsn	489
RESULT 4			
Q9DGR2		PRELIMINARY;	PRT; 767 AA.
ID	Q9DGR2		
AC	Q9DGR2		
DT	01-MAR-2001	(TEMBLrel. 16, Created)	
DT	01-MAR-2001	(TEMBLrel. 16, Last sequence update)	
DT	01-OCT-2003	(TEMBLrel. 25, Last annotation update)	
DE	Embryonic serine protease-2.		
DE	XISP-2.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
NCBI_TaxID=8355;			
GN	XISP-2.		
SE	SEQUENCE FROM N.A.		
MDLLINE=20363741; PubMed=10933452;			
RA	Yamada K., Takabatake T., Takeshima K.;		
RT	"-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.		
RT	from Xenopus laevis.;		
RT	Gene 252:209-216(2000).		
CC	ENEL; AB038497; BAB08217.1; -.		
DR	HSSP; P00766; 1CHK.		
DR	MEROPS; S01.049; -.		
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0004295; F:trypsin activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR009003; Cys Ser trypsin.		
DR	InterPro; IPR002172; LDL receptor A.		
DR	InterPro; IPR001254; Peptidase S1.		
DR	InterPro; IPR001314; Peptidase_S1A.		
DR	Pfam; PF000057; ldl_recept_a.4.		
DR	Pfam; PF000089; trypsin; 1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	PRINTS; PR00261; LDLRECEPTOR.		
DR	SMART; SM00192; LDLA; 8.		
DR	SMART; SM00020; Tryp SPC; 1.		
DR	PROSITE; PS01209; LDLRA_1; 8.		
DR	PROSITE; PS00068; LDLRA_2; 2.		
DR	PROSITE; PS02040; TRYPSIN_DOM; 1.		
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.		
DR	PROSITE; PS00135; TRYPSIN_SER; 1.		
KW	Hydrolase; Protease; Serine protease.		
SQ	SEQUENCE 767 AA; 86001 MW; E0566A38796D9E6E CRC64;		
Alignment Scores:			
Pred. No.:	4, 88e-80	Length:	767
Score:	1009.00	Matches:	189
Percent Similarity:	64.15%	Conservative:	49
Best Local Similarity:	50.94%	Mismatches:	131
Query Match:	31.74%	Indels:	2
DB:	13	Gaps:	2
US-09-323-597C-1 (1-1738) x Q9DGR2 (1-767)			
Qy	469	TGGGACTCTTCAGGTACTCGCATCAACCCCTCTACTGGTGTGATGGCGTGTCACTGC	528
Db	395	CysGlySerSerValSerCysValIleuSerSerGlnTrpCysAspGlyValSerAspCys	414
Qy	529	CCCGCGCGGGAGGACGAGAATCGGTGTGTTCGCCCTCTACGGAACCAACTTCACCTCTTCAG	588
Db	415	ProTyrGlyGluAspGluMetSerCysValSerLeuTyrProAlaAspPheGlnIleuGln	434
Qy	589	GTGTACTCATCTCAGAGGAAGTCCTGGCACCCTGTGTGCGCAAGACGACTGGGAACGAGAAC	648
Db	435	ValTyrSerThrValSerAlaTrpIleuProValCysSerAspTyrTrpAsnAspAsp	454

512 GlyValThrLysValThrGluValLeuProTyrPheTyrSerLysMetGluSerGlu 531

Db	512 GlyValThrLysValThrGluValLeuProTyrPheTyrSerLysMetGluSerGlu 531	QY	334	-----ACAGTGTGCACCTCAAAGACTAAAGAAAGCACTGTGCATC 372
Q9BYE2	PRELIMINARY; PRT; 581 AA.	Db	99	SerSerSerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeuValArg 118
Q9BYE2	Q9BYE2	QY	373	ACCTTGACCTGGGACCTTC----- 393
AC	Q9BYE2	Db	119	AlaThrProValGlyAlaValProIleArgSerSerProAlaArgSerAlaProAlaThr 138
DT	01-JUN-2001 (TRENBLrel. 17, Created)	QY	393	----- 393
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)	Db	139	ArgAlaThrArgGluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGln 158
DE	Membrane-type mosaic serine protease.	QY	394	-----CTCGTGGAGCT-----GCGTGGCGCGTGGCTACTC 426
GN	MSPL.	Db	159	LysGlnLeuProLeuIleGlyCysValLeuLeuLeuLeuAlaLeuValValSerLeuIle 178
OS	Homo sapiens (Human).	QY	427	-----TGAAGTTCATGGCAGCAGTCTCCAACTCTGGGATA--GAG 468
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Db	179	IleLeuPheGlnPheTrpGln-----GlyHisThrGlyIleArgHis 192
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	QY	469	TGCAGCTCTCAGGTACTGCATCAACCCCTTAACGTGTGTGATGCGTGTCACTGC 528
OX	NCBI_TaxID=9606;	Db	193	LysGlnArgGluSerCysProLysHisAlaValArgCysAspGlyValValAspCys 212
RN	(1)	QY	529	CCCGCGGGGAGGAGAGATCGGTGTGCTCTACGACCAAACTTCATCTTCAG 588
RP	SEQUENCE FROM N.A.	Db	213	LysLeuLysSerAspGluLeuGlyCysValArgPheAspTrpAspLysSerLeuLeuLys 232
RC	TISSUE=Lung; N.A.	QY	589	GTGTACTCATCTCAGAGAAAGTCTCTGTCACCTGTGTGCCAAGACGATGGAACGAGAAC 648
RX	MEDLINE=21167393; PubMed=11267681;	Db	233	IleTyrSerGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSer 252
RA	Kim D.R., Sharmin S., Inoue M., Kido H.;	QY	649	TACGGGGGGGGCGCTCAGGAGCATGGGTATAGAATAATTTTACTTAGCCAA---- 705
RT	"Cloning and expression of novel mosaic serine proteases with and	Db	253	TyrSerGluLysThrCysArgGlnLeuGlyPheGluSerAlaHisArgThrGluVal 272
RL	without a transmembrane domain from human lung.,"	QY	706	GGAATAGTGATGACAGCGATCCACAGCTTTATGAACCTTGAACAAAGTGGCGCAAT 765
DR	EMBL; AB048796; BAB39741.1; --	Db	273	AlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThr----- 289
DR	HSSP; P00783; IDPO.	QY	766	GTCGATATCTATAAAACTGTACACAGTGATCCCTGTCTTCTTCAAAGCAGTGTCT 825
DR	MEROFS; S01.087; --	Db	290	-----IleGlnGluSerLeuHisArgSerHisCysProSerGlnArgTyrIleSer 306
DR	GO; GO:0016021; C:integral to membrane; NAS.	QY	826	TTACGCTGTATAGCTCGGGGTCAACTTGAACCTCAAGCCGCGCAGACGAGATTGTGGC 885
DR	GO; GO:006508; P:proteolysis and peptidolysis; NAS.	Db	307	LeuGlnCysSerHisCysGlyLeuArg-----AlaMetThrGlyArgIleValGly 323
DR	InterPro; IPR009003; Cys Ser trypsin.	QY	886	GCGAGAGCGCGCTCCCGGGGGCTGCCCTGCGAGCTCAGCTGCAGTCAGAACGTC 945
DR	InterPro; IPR002172; LDL receptor A.	Db	324	GlyAlaLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThrThr 343
DR	InterPro; IPR001254; Peptidase S1.	QY	946	CACGTGTGGGAGGCTCCATCATCACCCTCCGAGTGGATCGTGACAGCGCCCATCGGTG 1005
DR	InterPro; IPR001314; Peptidase S1a.	Db	344	HisIleCysGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhe 363
DR	Pfam; PF00057; ldl_recept_a; 1.	QY	1006	-----GAAAAACCTCTTAACAATCCATGCGATTGGACGGCAATTGCGGGGATT 1053
DR	Pfam; PF00089; trypsin; 1.	Db	364	PheValThrArgGluLysValLeuGluGly-----TrpLysValTyrAlaGlyThr 380
DR	PRINTS; PR00722; CHYMOTRYPSIN.	QY	1054	TTGAGACAACTTTCATGTCTATGAGAGCGGATACCAAGTAGAGAAAAGTATTCTCAT 1113
DR	SMART; SM00192; LDLa; 1.	Db	381	SerAsnLeuHisGlnLeuProGluAlaAla-----SerIleAlaGluIleIleAsn 398
DR	SMART; SM00202; SR; 1.	QY	1114	CCAATTATGACTCCAAGACCAAGCAATGACATTCGGCTGATGAAGCTCAGAAGCT 1173
DR	PROSITE; PS0287; SRCK_2; 1.	Db	399	SerAsnTyrThrAspGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysPro 418
DR	PROSITE; PS0240; TRYPsin DOM; 1.	QY	1174	CTGACTTTCAACGACCTAGTGAACACGAGTGTGTCTGCCCAACCCAGGATGATCTGCAG 1233
DR	PROSITE; PS00134; TRYPsin HIS; 1.	Db	419	LeuThrLeuSerAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThrPheSer 438
DR	PROSITE; PS00135; TRYPsin SER; 1.	QY	1234	CCAGACAGCTCTCTCGATTTCGCGGTGGGGGGCCACCGAGGAG-----AAAGGAAGACC 1290
KW	Hydrolase; Protease; Serine protease.			
SQ	SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44 CRC64;			

Alignment Scores:

Pred. No.: 5.03e-56 Length: 581
 Score: 736.50 Matches: 177
 Percent Similarity: 46.67% Conservative: 82
 Best Local Similarity: 31.89% Mismatches: 197
 Query Match: 23.17% Indels: 99
 DB: 4 Gaps: 17

US-09-323-597c-1 (1-1738) x Q9BYE2 (1-581)

QY	124 TCAGGTCCACACCGATTTGGACTTACTATGAACCATGGATACCAACGGAAC 183
Db	27 AlaGlyThrProGlyArgAla-----SerProAlaGln 38
QY	184 CCTATCCGCGACAGCCCACTGFGCCCACT---GHCTACGAGGTGATCCGCTCAG 240
Db	39 AlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgAlaSerProAlaGln 58
QY	241 TACTACCGTCCCGGTGCC-----CAGTACGCGCCGAGGCTC 279
Db	59 AlaSerProAlaGlyThrProGlyArgAlaSerProGlyArgAlaSerProAlaGln 78
QY	280 CTGACGACGGCT---TCCAAACCGGTGCTCCACGACCCCAATCCCATCCGG--- 333
Db	79 AlaSerProAlaArgAlaSerProAlaLeuAlaSerLeuSerArgSerSerGlyArg 98

```

Db      439  LeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAspLysThr 458
QY      1291  TCAGAGTCTGAGACCTGCGCCAGAGTCTTCTCATTCAGACACAGAGATGCAACACGAGA 1350
Db      459  SerProPheLeuArgGluValGlnValAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 478
QY      1351  TATGCTATGACAACTGATACACACGAGCCATGATCTGTGCGGCTTCTGTCAGGGGAAC 1410
Db      479  LeuValTyAspSerTyLeuThrProArgMetMetCysAlaGlyAspLeuHisGlyGly 498
QY      1411  GTGATCTTCTGCGAGGTGACAGTGTGAGGCGCTCTGCTCATCTCGAAGACAATATCTCG 1470
Db      499  ArgAspSerCysGlnGlyAspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTrp 518
QY      1471  TGGCTATAGGCGATCAACAGCTGGGCTTCTGCTGTGCGCAACCTTACAGACAGAGTG 1530
Db      519  TyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyVal 538
QY      1531  TAGCGGAATGTGATGATTACGAGCTGAGTATTCGACAATG 1575
Db      539  TyrThrLysValThrGluValLeuProTrpIleTySerLysMet 553

RESULT 8
QSCFE0 PRELIMINARY; PRT; 471 AA.
AC QSCFE0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to mosaic serine protease (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EXBL; BC042878; AAH42878.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; P:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006506; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tlyp_SPC; 1.
DR PROSITE; PS00267; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
FT NON TER.
SQ SEQUENCE 471 AA; 52535 MW; ED58CF5B7C3BCC4 CRC64;

```

Alignment Scores:

```

Pred. No.: 1,45e-55 Length: 471
Score: 731.00 Matches: 166
Percent Similarity: 51.50% Conservative: 75
Best Local Similarity: 35.47% Mismatches: 183
Query Match: 22.99% Indels: 44
DB: 11 Gaps: 13

```

```

US-09-323-597C-1 (1-1738) x QSCFE0 (1-471)
QY      250  TCCCCGGTGCCCCAGTACGCCCGGAGGTCTCGACGAGGCTTCCAAACCCCGTCTCTGC 309
Db      19  SerProThrArgValTyLeuValArgAlaThrProValGlyAlaValProIleArgAla 38
QY      310  AGCAGCCCAATATCCCATCCCGGACAGTGTGCACC----- 345
Db      39  SerProAlaArgSerAlaProAlaThrArgAlaThrArgLysProGlyLeuSerPhe 58
QY      346  -----TCAAGACTAAGAAAGCACTG-----TGCATCACCTTG 378
Db      59  ProLysPheSerTrpGlnThrGlnArgGlnLeuProLeuLeuLeuLeuLeuLeuLeu 78
QY      379  ACCCTGGGACCTTCTCGTGGGACCTCGCTGGCGCTGGCGCTACTCTTGGAAAGTTCATG 438
Db      79  -----LeuLeuSerLeuValLeuLeuLeuLeuLeuLeuPheTyTrp 93
QY      439  GGCAGCAAGTGTCCAACTCTGGGATAGAGTGC-----GACTCTCTAGGTACCTGCATCAAC 495
Db      94  Arg-----GlyHisThrGlyIleLysTyLysGluProLeuGluSerCysProIle 110
QY      496  CCTCTTAACGTGTGTGGTGTGCACACTGCCCGCGGGGAGGACGAGATCGGTGT 555
Db      111  HisAlaValArgCysAspGlyValValAspCysLysMetLysSerAspGluLeuGlyCys 130
QY      556  GTTGCCTCTACGGACCAAACTTCATCTCTCAGGTGTACTCATCTCAGAGGAAGTCTGTGG 615
Db      131  ValArgPheAspTrpAspLysSerLeuLeuLysValTySerGlySerGlyGluTrp 150
QY      616  CACCTGTGTGTCGACAGACTGACAGCACTACGCGGCGGCGGCGCTGACAGGACATG 675
Db      151  LeuProValCysSerSerSerTrpAsnAspThrAspSerLysArgThrCysGlnGlnLeu 170
QY      676  GGTATAAGAATAATTTTACTCTAGCCAAGAAATAGTGGATGACAGCGGATCCACGAC 735
Db      171  GlyPhe---AspSerAlaTyArgThrThrGluValAlaHisArgAspIleThrSerSer 189
QY      736  TTTATGAACCTGACACAAAGTCCCGGCAATGTGATATCTATATAAAACTGTACACAGT 795
Db      190  PheLeu-----LeuSerGluTyAsnThrThrIleGlnGluSerLeuTyArgSer 206
QY      796  GATGCTCTTCTTCAAAAGCAGTGTGTTTCTTACCTGTATAGCTGTGCGGGGTCAACTTG 855
Db      207  Gln---CysProSerArgArgTyValSerLeuGlnCysSerHisCysGlyLeuArg--- 224
QY      856  AACTCAAGCCGCCAGACAGCAGATGTGTGGCGGCGAGAGCGGCTCCCGGGGCGCTGGCCC 915
Db      225  -----AlaMetThrGlyArgIleValGlyGlyAlaLeuThrSerGluSerLysTrpPro 242
QY      916  TGGCAGGTCCAGCTCCAGTCCAGAACGTCACAGTGTGCGGAGGCTCCATCATCACCCCC 975
Db      243  TrpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuLeuLeuAsp 262
QY      976  GAGTGGATCGTGACAGCGCCCACTGCGTG-----GAAAAACCTCTTTAAACAAT 1023
Db      263  GlnTrpValLeuThrAlaAlaHisCysPheValThrArgGluLysLeuLeuGly 282
QY      1024  CCATGGCATTCGACGGCATTTCCGGGGATTTGAGACAACTCTTTCATGTTCTATGAGCC 1083
Db      283  -----TrpLysValTyAlaGlyThrSerAsnLeuHisGlnLeuProGluAla 299
QY      1084  GGATACCAAGTAGAAAAAGTATTTCTCATCAAAATATGACTCCCAAGCAACCAAGAACAA 1143
Db      300  -----SerIleSerGlnIleIleAsnGlyAsnTyThrAspGluLeuAspAspTy 317
QY      1144  GACATTCGCTGATGACAGCTTCAGAGCTTCTGACTTTCAACGACCTAGTGAACCACTG 1203
Db      318  AspIleAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 337
QY      1204  TGTCTGCCCAACCCAGGCAATGATGTCGAGCCAGAACAGCTCTGCTGATTTCCGGGTGG 1263

```

```

Db      338 CysLeuProMetHisGlyGlnThrPheGlyLeuAsnGluThrCysTrpIleThrGlyPhe 357
QY      1264 GGGGCCACGACGAGAGAGGG---AAGACTCAGAGTCTGACGCTCCCAAGTGCTT 1320
Db      358 GlyLeuThrLysGluThrAspGluThrSerProPheLeuArgGluValGlnValAsn 377
QY      1321 CTCATTGACACACAGAGATGCAACACAGACAGATATGCTATGACAACTGATCACACAGCC 1380
Db      378 LeuIleAspPheLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArg 397
QY      1381 ATGATCTGTGCGGCTCTCTGCGGGGAAACGTCGATCTTTCAGGGTGACAGTGAGGG 1440
Db      398 MetMetCysAlaGlyAspLeuArgGlyIleArgSerCysGlnGlyAspSerGlyGly 417
QY      1441 CCTCTGCTCACTTCGAAGAACATATCTGCTGCTGATAGGAGGATACAGCTGGGGTCT 1500
Db      418 ProLeuValCysGluGlnAsnAsnArgTyrPheLeuAlaGlyValThrSerTrpGlyThr 437
QY      1501 GGCTGTGCGCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTCACGAGCTGG 1560
Db      438 GlyCysGlyGlnLysAsnLysProGlyValTyrThrLysValThrGluValLeuProTrp 457
QY      1561 ATTATTCGACAATGAGGCGACAC 1584
Db      458 IleTyrArgLysMetGluSerGlu 465

RESULT 9
Q86YM4
ID      Q86YM4      PRELIMINARY;      PRT;      558 AA.
AC      Q86YM4;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Transmembrane protease serine 6.
GN      TMPRSS6.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Park T.J., Park W.J.;
RT      "Homo sapiens transmembrane protease, serine 6 (TMPRSS6) mRNA.";
RL      Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY190317; AAO38062.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0004263; F:chymotrypsin activity; IEA.
DR      GO; GO:0008233; F:peptidase activity; IEA.
DR      GO; GO:0005044; F:scavenger receptor activity; IEA.
DR      GO; GO:0004295; F:trypsin activity; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      InterPro; IPR009003; Cys_Ser_trypsin.
DR      InterPro; IPR001254; Peptidase_S1.
DR      InterPro; IPR001314; Peptidase_S1A.
DR      InterPro; IPR001190; Srcr_receptor.
DR      Pfam; PF00089; trypsin; 1.
DR      PRINTS; PR00722; CHYMOTRYPSIN.
DR      SMART; SM00202; SR; 1.
DR      SMART; SM00020; Tryp_SPC; 1.
DR      PROSITE; PS0287; SRCR_2; 1.
DR      PROSITE; PS0240; TRYPSIN_DOM; 1.
DR      PROSITE; PS00134; TRYPSIN_HIS; 1.
DR      PROSITE; PS00135; TRYPSIN_SER; 1.
KW      Protease.
SQ      SEQUENCE 558 AA; 60432 MW; 4AC817FCD70D7017 CRC64;

```

```

Alignment Scores:
Pred. No.:      4,658-55      Length:      558
Score:          725.50      Matches:    174
Percent Similarity: 45.39%      Conservative: 82
Best Local Similarity: 30.85%      Mismatches: 195
Query Match:      22.82%      Indels:     113
Db:              4          Gaps:      16

```

US-09-323-597C-1 (1-1738) x Q86YM4 (1-558)

```

QY      124 TCAGGCTCACCAACAGCTATTGGACCTTACTATGAAAACCATGATACCAACCGAACAAC 183
Db      27 AlaGlyThrProGlyArgAla-----SerProAlaGln 38
QY      184 CCCTATCCGCGACACCCACCTGTGTGTCCTCCCACT---GTCTACGAGGTGCATCCGGCTCAG 240
Db      39 AlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgAlaSerProAlaGln 58
QY      241 TACTACCGCTCCCGCGTGCCTC-----CAGTACGCCCGCAGGGGTC 279
Db      59 AlaSerProAlaGlyThrProGlyArgAlaSerProGlyArgAlaSerProAlaGln 78
QY      280 CTGACGAGGCT---TCAACCCCGTGTCTGCGACGAGCCCAAAATCCCTCCCGG--- 333
Db      79 AlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgAlaSerProGlyArg 98
QY      334 -----ACAGTGTGCACTCAAAGACTAAAGACACTGTCGATC 372
Db      99 SerSerSerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeuValArg 118
QY      373 ACCTTGACCTCGGGGACCTTC----- 393
Db      119 AlaThrProValGlyAlaValProIleArgSerSerProAlaArgSerAlaProAlaThr 138
QY      393 ----- 393
Db      139 ArgAlaThrArgGluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGln 158
QY      394 -----CTGCTGGAGCTGCGCTGGCC----- 414
Db      159 LysGlnLeuProLeuIleGlyCysValLeuLeuLeuAlaLeuValValSerLeuIle 178
QY      415 -----GCTGGCTACTCTCGAAGTTCATGGGACGACG 447
Db      179 IleLeuPheGlnPheTrpGlnGlyTyrThrGlyIleArgTyrLysGluGlnArgGluSer 198
QY      448 TGCTCCCAACTCTGGGATAGAGTGGGACTCCTCAGGTACTGTCATCAACCCCTCTAACTGG 507
Db      199 CysProGluHisAlaValArg----- 205
QY      508 TGTGATGGCGTGTCACTCCCGCGGGGAGGACGAGATCGGTGTGTGGCTCTTAC 567
Db      206 ArgAspGlyValValAspCysLysLeuLysSerAspGluLeuGlyCysValArgPheAsp 225
QY      568 GGACCAAACTTCATCTTCAGGTGTCTCATCTCAGAGGAAGTCTCGCACCTCTGTGTC 627
Db      226 TrpAspLysSerLeuLeuLysIleTyrSerGlySerSerHisGlnTrpLeuProIleCys 245
QY      628 CAAGACGACTGGAACGAGAACTACGGCGGGCGGCTGCGAGGACATGGGCTATAAGAAT 687
Db      246 SerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSer 265
QY      688 AATTTTACTCTAGCCAA---GGAAATAGTGTGATGACAGCGGATCCACAGCTTTATGAAA 744
Db      266 AlaHisArgThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArg 285
QY      745 CTGAACACAAAGTGCCTGCAATGTCTGATATCTATAAAACTGTACACAGATGTCCTGT 804
Db      286 TyrAsnSerThr-----IleGlnGluSerLeuHisArgSerGlu---Cys 299
QY      805 TCTTCAAAAGCAGTGGTTTCTTTACGCTGTATAGCTCGCGGGTCACTTGAACCTCAAGC 864
Db      300 ProSerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLeuArg-----Ala 316
QY      865 CGCCAGACGACGAGATTGTGGCGCGGAGAGCGCGCTCCCGGGGGCGCTCGCCCTGCGAGTC 924
Db      317 MetThrGlyArgIleValGlyAlaLeuAlaSerAspSerLysTrpProTrpGlnVal 336
QY      925 AGCTGCACTCCAGAACGTCACGTTGCGGAGGCTCCATCATCCCCCGGAGTGGATC 984

```

```
Db 337 SerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuLeuAspAlaGlnTrpVal 356
Qy 985 GTGACGCCGCCCTCGTGC-----GAAAAAAGCTCTTAACAATCCATGCCAT 1032
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 357 LeuThrAlaAlaHisCysPhePheValThrArgGluLysValLeuGluGly----- 373
Qy 1033 TGGACGCCATTGGGGGATTTGAGACAATCTTTCATGTTCTCTAGGAGCGGATACAA 1092
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 374 TrpLysValTyAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAla-----Ser 391
Qy 1093 GTAGAAAAGATGTTCTCTCAATATGACTCCAGACACAGAACCAATGACATTCGCG 1152
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 392 IleAlaGluIleIleAsnSerAsnTyThrAspGluGluAspTyThrAspIleAla 411
Qy 1153 CTGATGAAGCTGCAGAGCTCTGACTTCAACAGCTAGTGAACCAAGTGTGTCTCC 1212
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 412 LeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleHisProAlaCysLeuPro 431
Qy 1213 AACCCAGGCATGATGCTGCAGCCAGACAGCTCTGCTGGATTTCCGGTGGGGCCACC 1272
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 432 MethHisGlyGlnThrPheSerLeuAsnGluThrCysTrpIleHrnGlyPheGlyLysThr 451
Qy 1273 GAGGAG---AAAGGGAAGACCTCAGAAGTGTGTAACGCTGCCAAGTGTCTTCTCAT 1329
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 452 ArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuLeuAsp 471
Qy 1330 ACACAGAGATGCACACAGATATGCTGATCAGCAACCTGATCACCAGCCATGATCTGT 1389
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 472 PheLysCysAsnAspTyLeuValTyAspSerTyLeuThrProArgMetMetCys 491
Qy 1390 GCCGGCTTCCTGCAGGGGAAAGCTCGATTCTGCGAGGCTGACAGTGGAGGCTCTGCTC 1449
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 492 AlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyProLeuVal 511
Qy 1450 ACTTCGAAGAACATATCTGTGCTGATAGGGAATACAGCTGGGTTCTGCTGTGCC 1509
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 512 CysGluGlnAsnAsnArgTrpTyLeuAlaGlyValThrSerTrpGlyThrGlyCysGly 531
Qy 1510 AAGCTTACAGACACAGAGTGTGCGGAATGCTGATGCTATTCAGGACTGGATTTATCGA 1569
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 532 GlnArgAsnLysProGlyValTyThrLysValThrLysValLeuProTrpIleTySer 551
Qy 1570 CAATGAGGGCA 1581
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 LysMetGluSer 555
RESULT 10
Q96B86 PRELIMINARY; PRT; 405 AA.
AC Q96B86;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strauberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -.
DR HSSP; P00761; 1AN1.
DR MEROPS; S01.034; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0004283; F: chymotrypsin activity; IEA.
DR GO; GO:0008233; F: peptidase activity; IEA.
DR GO; GO:0005044; F: scavenger receptor activity; IEA.
DR GO; GO:0004295; F: trypsin activity; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
```

```
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1_
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM0020; Tryp_SPC; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON_TER 1 1
SQ SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;

Alignment Scores: 8.97e-51 Length: 405
Pred. No.: 676.50 Matches: 150
Score: 676.50
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 21.28% Indels: 49
DB: Gaps: 13

US-09-323-597C-1 (1-1738) x Q96B86 (1-405)
Qy 508 TGTGATGGGTGTGCACACTGCCCCGGGGAGGAGAGAAATCGTGTGTTCGCTCTAC 567
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 45 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 64
Qy 568 -----GGACCA-----AACTTCATCTCTCAGGTGTAC 594
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 84
Qy 595 TCATCTCAGAGGAAGTCCTGGCACCTGTGTGTCACAGACAGCAGCGAACGAGAACTACGGG 654
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 104
Qy 655 CGGCGCGCTCGCAGGACATGGCTAT-----AGAAATAATTTTACTTACGCCAA--- 705
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 GluThrAlaCysArgGlnMetGlyTySerSerLysProThrPheArgAlaValGluIle 124
Qy 706 -----GGAATAGTGGATGACGCGGATCCACCGACTTATGAACCTG 747
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 144
Qy 748 AACACAAGTCCCGGCAATGTCGATATCTATAAAAAAAGTGTACCATGATGCTCTTCT 807
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 ArgAsnSerSerGly-----ProCysLeu 152
Qy 808 TCNAAAGCAGTGGTTTCTTTTACGTGTATAGCTGCGGGGTCAACTTGAAGTCAAGCCGC 867
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 153 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 169
Qy 868 CAGAGCAGGATGTGGCGCGCAGAGCGCGCTCCCGGGCGCTGGCCCTGAGGAGTGCAGC 927
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 ThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSer 189
Qy 928 CTGCAGCTCCAGAACTGTCAGTGTGCGGAGCTCATCATCATCCCCCGAGTGGATCGTG 987
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 IleGlnTyAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 209
Qy 988 ACAGCGCCCACTGGCTGGAAAAAAGCTCTTAACAATCCATGCGATGGAGCGCATTTGCG 1047
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 210 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 228
Qy 1048 GCGATTTTGAGA---CAATCTTTTTCATGTTCTATGGAGCGGATACCAAGTAGAAAAAGTG 1104
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 229 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 244
```

```

QY 1105 ATT-----TCTCATCCAAATATGACTCCAGACCCAGAACCAATGACATTCGGCTG 1155
Db 245 lilellelleGluPheAsnProMetTyr |||||:|||||:|||||:|||||:|||||
QY 1156 ATGAAGCTGCAGAACCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTGCGCCAAAC 1215
Db 262 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 281
QY 1216 CCAGGCATGATCTCCAGCCAGAACAGCTCTGCTGGATTTCGGGTGGGGGGCCACCCAG 1275
Db 282 PheAspGluGluLeuThrProIleThrProLeuThrPheIleGlyTrpGlyPheThrLys 301
QY 1276 GAGAAA---GGAGAACCTCAGAACTGCTGAAGTGTGAACCTGCCAAGGTCTTCTCATTCAGACA 1332
Db 302 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 321
QY 1333 CAGAGATCCACAGCAGATATCTCTATGACAACTGATCAGACACGACCATGATCTGTGCC 1392
Db 322 ThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAla 341
QY 1393 GGCCTTCCTCCAGGGGAACGTCCATTCTTCCAGGCTGACAGTGGAGGCGCTCTGCTCACT 1452
Db 342 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 361
QY 1453 TCGAAGAACATATCTGCTGCTGATAGCGGATACAGCTGGGTTCGGCTGCGCCAAA 1512
Db 362 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 380
QY 1513 GCTTACAGACAGGAGTGTACGGAACTGTGATGTGATTCACGAGCTGATTCAGCAA 1572
Db 381 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 400
QY 1573 ATGAGGGCAGAC 1584
Db 401 TrpLysAlaGlu 404

RESULT 11
Q8CUJ7
ID Q8CUJ7 PRELIMINARY; PR7; 445 AA.
AC Q8CUJ7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenal mitochondrial protease long variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEDH;
RA Omer S.; Bicknell A.B.; Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EXBL; AF37098; AAN06757.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

```

```

KW Protease.
SQ SEQUENCE 445 AA; 48440 MW; BE3P56D8372ED988 CRC64;

Alignment Scores:
Pred. No.: 9,21e-51 Length: 445
Score: 676.50 Matches: 153
Percent Similarity: 48.67% Conservative: 66
Best Local Similarity: 34.00% Mismatches: 195
Query Match: 21.28% Indels: 36
DB: 11 Gaps: 9

US-09-323-597C-1 (1-1738) x Q8CUJ7 (1-445)
QY 273 GAGGTCCTCAGCAGCGCTTCCAAACCCGCTGCTGACGACGCCAAATCCCATCCGG 332
Db 8 GllGlyProGlyProGlyIlePheArgAlaGluLeuGlyAspGlnGlnProIleSer 27
QY 333 GACAGTGTGCACCTCAAAGACTAAGAAAGCACTGTGCATCACCTTGCACCTCGGGACCTT 392
Db 28 GlnSerGlnArgTrp-CysCysLeuGlnArgGlyCysValIleleuGlyAlaLeuGlyLe 47
QY 393 CTTCTGGGAGCTGCGCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452
Db 47 uLeuAlaGlyAlaGlyValGlySerTrpLeuLeuValLeuTyrLeuTrpProAlaAlaSe 67
QY 453 CAACCTCT---GGGATAGAGTGCACCTCCTCAGGTACTGTCATCAACCCCTCTAACTGGT 509
Db 67 rProProValSerValThrLeuGlnGluGluValThrLeuSer----- 82
QY 510 TGATGGGTGTACACTGCTGCGCGGGGAGGAGAGATCGGTGTGTT----- 558
Db 83 -----CysProGlyValSerSerGluGluLysLeuLeuProSerLeuPr 97
QY 559 -----CGCTCTACGACCAAACTTCATCTTCAGGTGTACTCATCTCA 602
Db 97 oLysAlaValSerPheArgIleAsnGlyGluAspLeuLeuGluValGlnValArgAl 117
QY 603 GAGGAAGTCTGCGACCCCTGTGTGCGCAAGACGATCGAAGAACTACGGCGGGCGGC 662
Db 117 aArgProAspTrpLeuLeuValCysHisGluGlyTrpAsnProAlaLeuGlyMethHisI 137
QY 663 CTGACGAGCATGGGTCTAAAGATAATTTTACTCTAGCCAA----- 705
Db 137 eCysGlnSerLeuGlyTyr-----PheArgLeuThrGlnHisLysAlaValAsnLe 154
QY 706 -GGAATAGTGTGATGACAGCGGATCCACCAGCTTTATGAACATGAACAGTCCGCGCAA 764
Db 154 uSerAspIleLysLeuAsnArgSerGlnGluPheAlaGlnLeuSerAlaArgProGlySe 174
QY 765 TGTGATATCTATAAAAAAAGTGTACACAGTGTGCTTCTTCAAAAGCAGTGTGTTTC 824
Db 174 rLeu----ValGluGluAlaTrpGlnProSerThrAsnCysProSerGlyArgIleValSe 193
QY 825 TTTACGCTGTATAGCTCGGGGTCAACTTGAACCTCAAGCCGCGAGCAGCAGGATTCGGG 884
Db 193 rLeuLysCysSerGluCysGlyAlaArg-----ProLeuAlaSerArgIleValGI 210
QY 885 CGGCGAGCGCGCTCCCGGGGCTGCGCTGCGGAGTCCAGCTGCGACGTCCAGAACGT 944
Db 210 yGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerValMetLeuGlySerAr 230
QY 945 CCAAGTGTGCGGAGGCTCCATCATCACCCCGAGTGGATCGTGAAGCCGCCACCTGCGT 1004
Db 230 gHisThrCysGlyGlySerValLeuAlaProTyrTrpValValThrAlaHisCysMe 250
QY 1005 GGAAAAA---CCTCTTAACAATCCCATGGCATTTGGACGGCATTTGGGGATTTTGAGACA 1061
Db 250 tTyrSerPheArgLeuSerArgLeuSerSerTrpArgValHisAlaGlyLeuValSerHi 270
QY 1062 ATCTTTCATGTTCTATGGAGCGGATACCAAGTGTGAAGAGTATTTCTCATCCCAATTA 1121
Db 270 sSerAlaValArgGlnHisGlnGlyThrMetValGluLysIleIleProHisProLeuTy 290

```



```
QY 1122 TGACTCCAGACCAAGACATGATTCGCTGATGAGCTGAGAGCTGCGCTGACTTT 1181
Db 290 rSerAlaGlnAenHisAspTyrAspValAlaLeuGlnLeuArgThrProIleAenPh 310
QY 1182 CAACGACCTAGTGAACACAGTGTCTGCCCCAACCCAGGATGATGCTGCAGCCAGACA 1241
Db 310 eSerAspThrValSerAlaValCysLeuProAlaLysGlnHisPheProGlnGlySe 330
QY 1242 GCTCTGCTGATTTCCGGTGGGGGCCACCGAG---GAGAAAGGAGACCTCAGAAGT 1298
Db 330 rGlnCysTrpValSerGlyTrpGlyHisThrAspProSerHisThrHisSerAspTh 350
QY 1299 GCTGAACGCTGCCAGGTGCTTCATTCATGACAGACAGAGTCAACAGCAGATATGCTA 1358
Db 350 rLeuGlnAspThrMetValProLeuLeuSerThrAspLeuCysAsnSerSerCysMetTy 370
QY 1359 TGACAACTGATCACACAGCCATGATCTGCGCGGCTTCCTCAGGGGAACTGCGATTC 1418
Db 370 rSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAspGlyArgAlaAspAl 390
QY 1419 TTGCAGGGTGCACAGTGGAGGCTCTGCTCCTCAGTTCGAGAACATATCTGCTGCTGAT 1478
Db 390 aCysGlnGlyAspSerGlyProLeuValCysProSerGlyAspThrTrpHisLeuVa 410
QY 1479 AGGGATACAGCTGGGGTCTGCTGCTGCGAAAGCTTACAGACGAGAGTGTACGGGAA 1538
Db 410 lGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgProGlyValTyrAlaLy 430
QY 1539 TGTGATGTTATTCAGCACTGGATTAT 1566
Db 430 sValAlaGluPheLeuAspTrpIleHis 439

RESULT 12
Q8CDRO PRELIMINARY; PRT; 455 AA.
AC Q8CDRO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transmembrane protease.
GN TWPRSS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
DR EMBL; AK029714; BAC26577.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS02287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 455 AA; 49669 MW; BE22B2E7503C74B CRC64;
```

```
Alignment Scores:
Pred. No.: 1,06e-49 Length: 455
Score: 664.50 Matches: 145
Percent Similarity: 50.97% Conservative: 66
Best Local Similarity: 35.02% Mismatches: 176
Query Match: 20.90% Indels: 27
DB: 11 Gaps: 9

US-09-323-597C-1 (1-1738) x Q8CDRO (1-455)
QY 367 TGCATCACCTTGACCTCGGGGACCTTCCTCGTGGAGCTGCGTGGCGCTGCGCTACTC 426
Db 49 CysValIleLeuGlyValLeuGlyLeuLeuAlaGlyAlaGlyIleAlaSerTrpLeuLeu 68
QY 427 TGAAGTTTCATGGCGAGCAAGTCTCTCAACTCTGGGATAGAGTGGAGTCTCTCAGGTACC 486
Db 69 ValLeuTyrLeuTrpProAlaSerProSer-----IleSerGlyThr 83
QY 487 TGCATCAACCCCTCACTAAGTGGTGTGATGGCGTGTACACTGCGCGCGGGAGGAGGAG 546
Db 84 LeuGlnGluGlu-----GluMetThrLeuAsnCysProGlyValSerArgGlu 99
QY 547 AATCGGTGTGTT-----CGCCTCTACGACCAAACTTC 579
Db 100 GluGluLeuLeuProSerLeuProLysThrValSerPheArgIleAsnGlyGluAspLeu 119
QY 580 ATCCTTCAGGTGACTCATCTCAGAGAAAGTCTCGGACCCCTGTGTGCCAAGAGACTGG 639
Db 120 LeuLeuGlnValGlnValAlaArgProAspTrpLeuLeuValCysHisGluGlyTrp 139
QY 640 AACGAGAACTACGGCGGGCGGCTGCAGGAGATGGGCTAT---AAGATAATTTTTC 696
Db 140 SerProAlaLeuGlyMetHisIleCysLysSerLeuGlyHisIleArgLeuThrGlnHis 159
QY 697 TCTAGCCAAAGTAATAGTGATGAC-----AGCGGATCCACAGCTTTATGAACCTGAAC 750
Db 160 LysAlaValAsnLeuSerAspIleLysLeuAsnArgSerGlnGluPheAlaGlnLeuSer 179
QY 751 ACAAGTGGCGCAATCTCGATATCATATAAACTGTACCAGAGTACGCTGCTTCTTCA 810
Db 180 AlaArgProGlyGlyLeu---ValGluGluSerTrpLysProSerSerAlaAsnCysProSer 198
QY 811 AAGCAGTGGTTTCTTTACGTGTATAGCTCGGGGCTCAACTTGAACCTCAAGCCGCCAG 870
Db 199 GlyArgIleValSerLeuLysCysSerGluCysGlyAlaArg-----ProLeuAla 215
QY 871 AGCAGGATTGCGCGCGGAGAGCGGCTCCCGGGGCGCTCCCGGGGCGCTGCGAGCTCAGCCTG 930
Db 216 SerArgIleValGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerVal 235
QY 931 CACGTCCAGAACTCCACGTCGCGAGGCTCCATCATCACCCCGAGTGATCGTGACA 990
Db 236 MetLeuGlySerArgHisThrCysGlyAlaSerValLeuAlaProHisTrpValValThr 255
QY 991 GCGGCCACTGCGTGGAATAA---CCTTTAAACAATCCATGGCATTTGGACGGCATTTGGG 1047
Db 256 AlaAlaHisCysMetTyrSerPheArgLeuSerArgLeuSerSerTrpArgValHisAla 275
QY 1048 GGGATTTTGAGACAATCTTTTCATGTTCTATGAGCGGATCCACAGTACAAAAGTGATT 1107
Db 276 GlyLeuValSerHisGlyAlaValArgGlnHisGlnGlyThrMetValGluLysIleIle 295
QY 1108 TCTCATCCAAATATGACTCCAAAGACCAAGAACATGACATTCGCGTGTAGAGCTGCAG 1167
Db 296 ProHisProLeuTyrSerAlaGlnAsnHisAspTyrAspValAlaLeuLeuGlnLeuArg 315
QY 1168 AAGCCTCTGACTTTCACAGCCTAGTGAACACGAGTGTCTGCCCAACCCAGGCATGATG 1227
Db 316 ThrProIleAsnPheSerAspThrValGlyAlaValCysLeuProAlaLysGluGlnHis 335
QY 1228 CTCGAGCCAGAACAGCTCTCTGTGGATTTCGGGTGGGGGGCCACCGAG---GAGAAAGG 1284
```

```
Db 336 PheProTrpGlySerClnCysTrpValSerGlyTrpGlyHisThrAspProSerHisThr 355
QY 1285 ANGACCTCAGAGTCTGCAAGCTGCAAGGTGCTTCTCATTTGAGACACAGAGATGCAAC 1344
Db 356 HisSerSerAspThrLeuGlnAspThrMetValProLeuLeuSerThrTyrLeuCysAsn 375
QY 1345 AGCAGATATGCTATGACAACTGATCACCAGCATGATCTGTGCGGCTTCCTGCAG 1404
Db 376 SerSerCysMetTyrSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAsp 1404
QY 1405 GGAACCTGATCTTCCAGAGGTGACAGTGGAGGCTCTGCTCACTTCGAGAACAAAT 1464
Db 396 GlyArgAlaAspAlaCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAsp 415
QY 1465 ATCTGTGCTCATAGGAGATCAAGCTGGGCTTCTGGCTGTGCCAAAGCTTACAGACCA 1524
Db 416 ThrTrpHisLeuValGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgPro 435
QY 1525 GGAGTGTCCGGAATGTATGTATTCACGAGTCTGATTTAT 1566
Db 436 GlyValTyrAlaLysValAlaGluPheLeuAspTrpIleHis 449

RESULT 13
ID Q8CJ16 PRELIMINARY; PRT; 371 AA.
AC Q8CJ16;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-NOV-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEOH;
RA Omer S.; Bicknell A.B.; Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF537099; AAN06758.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Strc_receptor.
DR Pfam; PF00089; trypsin; I.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;
```

```
Alignment Scores:
Pred. No.: 1,51e-49 Length: 371
Score: 662.50 Matches: 133
Percent Similarity: 53.82% Conservative: 57
Best Local Similarity: 37.68% Mismatches: 148
Query Match: 20.84% Indels: 15
DB: 11 Gaps: 6

US-09-323-597C-1 (1-1738) x Q8CJ16 (1-371)
```

QY 532 GCGCGGGAGGACGAGATCGGTGTGTCCTCTACGGACCAACTTCATCCTTCAGGTG 591

```
Db 20 GlyAspGlnGlnGlnProIleSerPheArgIleAsnGlyGluAspLeuLeuGluVal 39
QY 592 TACTCATCTCAGAGGAAGTCTCGCACCTGTGTGCCAAGACGACTGGAACAGCAACTAC 651
Db 40 GlnValArgAlaAArgProAspTrpLeuLeuValCysHisGlnGlyTrpAsnProAlaLeu 59
QY 652 GGGCGGGCGGCTCAGGAGCATCGGCTATAAGATAATTTTACTCTAGCCAA----- 705
Db 60 GlyMetHisIleCysGlnSerLeuGlyTyr-----PheArgLeuThrGlnHisLys 76
QY 706 -----GGAATAGTGGATGACAGCGGGATCCACCACTTTATGAACCTGAACACA 753
Db 77 AlaValAsnLeuSerAspIleLysLeuAsnArgSerGlnGluPheAlaGlnSerAla 96
QY 754 AGTCCCGCAATGTCGATATCTATAAAACTGTACACAGTGTGCTGCTCTTCAAAA 813
Db 97 ArgProGlySerLeu--ValGluGluAlaTrpGlnProSerThrAsnCysProSerGly 115
QY 814 CGAGTGGTTCTTTTACGCTGTATAGCTCGGGGTCACTTGAACCTCAAGCCGCGCAGAGC 873
Db 116 ArgIleValSerLeuLysCysSerGluCysGlyAlaArg-----ProLeuAlaSer 132
QY 874 AGGATTGCGGCGCGAGAGCGGCTCCCGGGGCTGGCCCTGGCAGGTGAGCTGCAC 933
Db 133 ArgIleValGlyGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerValMet 152
QY 934 GTCCAGAACCTCCAGCTGCGGAGGCTCATCATCACCCCGAGTGTGATCGTGCACAGCC 993
Db 153 LeuGlySerArgHisThrCysGlyGlySerValLeuAlaProTyrTrpValValThrAla 172
QY 994 GCCCACTGCTGGGAAAAA---CCTTTAAACAATCCATGGCATTTGGACGGCATTTGCGGG 1050
Db 173 AlaHisCysMetTyrSerPheArgLeuSerArgLeuSerSerTrpArgValHisAlaGly 192
QY 1051 ATTTTGAGACAACTCTTTCATGTTCTATGGAGCGGATACCAAGTAGAAAAAGTATTCT 1110
Db 193 LeuValSerHisSerAlaValArgGlnHisGlnGlyThrMetValGluLysIleIlePro 212
QY 1111 CATCAAAATTATGATCTCCAGACCCAGAACAAATGACATTCGCTGATGAAAGTGCAGAG 1170
Db 213 HisProLeuTyrSerAlaGlnAsnHisAspTyrAspValAlaLeuLeuGlnLeuArgThr 232
QY 1171 CCTCTGACTTTCAACGACCTAGTGAACACAGTGTGTCTGCCCAACCCAGGCATCATCTG 1230
Db 233 ProIleAsnPheSerAspThrValSerAlaValCysLeuProAlaLysGluGlnHisPhe 252
QY 1231 CAGCCAGAACAGCTCTGCTGGATTTCGGGTGGGGGGGCCACCGAG---GAGAAAGGGAAG 1287
Db 253 ProGlnGlySerGlnCysTrpValSerGlyTrpGlyHisThrAspProSerHisThrHis 272
QY 1288 ACCTCAGAGTGTGTAACCGTCCAGAGTGTCTTCTCATTTGAGACACAGATGCAACAGC 1347
Db 273 SerSerAspThrLeuGlnAspThrMetValProLeuLeuSerThrAspLeuCysAsnSer 292
QY 1348 AGATATGCTCTATGACAACTGATCACCAGCCATGATCTGTCCGGCTTCTCTGCGAGGG 1407
Db 293 SerCysMetTyrSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAspGly 312
QY 1408 AACGTGATTCTTGGCCAGGGTGTACAGTGGAGGGCTCTGTGCTACTCTCGAAGAACATATC 1467
Db 313 ArgAlaAspAlaCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAspThr 332
QY 1468 TGTGTGCTGATAGGGGATACAACTGGGTTCCTGGCTGTGCCCAAGCTTACAGACAGGA 1527
Db 333 TrpHisLeuValGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgProGly 352
QY 1528 GTCTACGGGAATGTGATGTATTCACGAGCTGATTTAT 1566
Db 353 ValTyrAlaLysValAlaGluPheLeuAspTrpIleHis 365

RESULT 14
Q7Z280
```

```

ID Q7Z80 PRELIMINARY; PRT; 326 AA.
AC Q7Z80;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI-d269G10.3 (Novel protein similar to human transmembrane protease,
DE serine 3 (TPRSS3)) (Fragment).
GN SI-d269G10.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL672083; CAD61105.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0004263; F: chymotrypsin activity; IEA.
DR GO; GO:0008233; F: peptidase activity; IEA.
DR GO; GO:0005044; F: scavenger receptor activity; IEA.
DR GO; GO:0004295; F: trypsin activity; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR02172; LDL receptor A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PS00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Transmembrane; 1
FT NON_TER
SQ SEQUENCE 326 AA; 35561 MW; 8D6F12214393CDB2 CRC64;

Alignment Scores:
Pred. No.: 2,26e-48 Length: 326
Score: 649.00 Matches: 240
Percent Similarity: 50.13% Conservative: 50
Best Local Similarity: 36.94% Mismatches: 121
Query Match: 20.42% Indels: 68
DB: 13 Gaps: 8

US-09-323-597C-1 (1-1738) x Q7Z80 (1-326)
QY 439 GCGAGCAGTCTCAACTGGGATAGAGTGGACTCTCAGTACTGCATCAACCC 498
Db 1 GlyLeuArgSerCysSerGlyLysPheHisCysValSerValArgCysileSerArg 20
QY 499 TCTAACTGGTGTGATGGGTCTCACATGCCCGCGGAGGACGAGATCGGTGT 558
Db 21 AsnAlaValCysAspGlyValGlnAspCysArgAspGlyGluAspGluLeuAsnCysVal 40
QY 559 CGCTCTACGACCAAACTCATCTCAGTGTACTCATCTCAGAGAAAGTCTCGCAC 618
Db 41 ArgValSerGlySerHisSerValLeuGlnValPheGly-----ArgGlyLeuTrpArg 58
QY 619 CCTGTGTCACAGACGACTGCAACGAGACTACGGCGGGCGGCTCGAGGACATGGC 678
Db 59 ThrValCysSerGluGlyTrpAspSerGlnLeuSerThrLeuAlaCysArgGlnLeuGly 78
QY 679 TATAAGAATAATTTTACTCTAGCCAAAGGAATAGTGGATGACAGCGGATCCACGCTTT 738

```

```

79 Tyr----- 79
QY 739 ATGAAGACTGAACACAGTGGCGCAATGTGCGATATCTATAAAACTGTACCACAGTGT 798
79 ----- 79
QY 799 GCCTGTCTTCAAAAGCAGTGGTTCCTTTAGCGTGTATAGCTGCGGGGTCAACTTGAC 858
80 -----SerArgSerAlaIleSer----- 85
QY 859 TCAAGCGCGCAGACGACGATGTGGGGCGGAGAGCGCGCTCCCGGGGGCTGCGCCCTGG 918
86 -----SerArgIleValGlyGlyAsnValSerLysSerGlyGlnValProTrp 101
QY 919 CAGTCAGCTGCAGCTCCAGAGCTCCAGTGTGCGGAGGCTCCATCATCACCCTCGAG 978
102 GlnValSerLeuHisTyrGlnAsnGlnTyrLeuCysGlyGlySerIleIleSerGluSer 121
QY 979 TGGATCTGACAGCGCCACTGCTGGTGGAAAAACCTCTTAACAATCCATCGCATTCAGC 1038
122 TrpIleLeuThrAlaAlaHisCysVal----PheGlyPheAlaGlnProValLeuTrpAsp 140
QY 1039 GCATTTTCGGGGATTTTGACACAATCTTTCATGTTCTATGAGCGCGGATACCAAGTAGAA 1098
141 ValTyrAlaGlyLeuIleAsnLeuPro----LeuSerLysAlaGluAlaHisSerValGlu 159
QY 1099 AAGTGTATTTCTCATCCAAATTATGACTCCAGACCAAGAACATGATGATGATGCTGCTG 1158
160 LysIleIleTyrHisAlaAsnPheArgSerLysSerPheSerTyrAspIleAlaLeuIle 179
QY 1159 AAGCTGCAAGAGCTCTGACTTTCAACGACTAGTGAACAGTGTCTTCTGCGCCCAACCCA 1218
180 LysLeuThrLeuProLeuThrPheAsnAspGlnIleAlaProIleCysLeuProAsnTyr 199
QY 1219 GGCATGATGTCGACCCAGACAGCTCTGCTGATTTCCGGTGGGGGGCCACGAGGAG 1278
200 GlyGluSerPheLysAsnGlyGlnMetCysLeuIleSerGlyTrpGlyAlaThrValAsp 219
QY 1279 AAGAGGAAGACCTCAGAGTGTGCAAGCTGCGCAAGTGTCTTCATTGAGACACAGAGA 1338
220 SerGlyGluThrSerLeuSerLeuHisValAlaGlnValProLeuLeuSerAsnLysGlu 239
QY 1339 TGCAACAGCAGATATGTTCTATGACCACTGATCACACCGCCATGATCTGTGCGGGTTC 1398
240 CysArgLysLeuGly-----LeuThrAsnTrpAsnValCysThrGluPhe 254
QY 1399 CTGCAGGGGAACCTGCGATTCTTCCAGGGTGACAGTGGAGGGCTCTGTCACCTCGAAG 1458
255 LeuArgGly---ValGlyThrCysGlnGlyAspSerGlyGlyProLeu---AlaCysGln 272
QY 1459 AACAAATATCTGCTGCTGATAGGGGATACAAAGCTGGGGTTCGCTGTGCCAAGCTTAC 1518
273 GlySerAlaTrpThrLeuValGlyThrGlySerTrpAspGluAsnCysGlyLysValAsn 292
QY 1519 AGACCGAGGAGTGTACGGGAATGTGATGTGATTACCGACTGGATTTATCGACAAATG 1575
293 LysProGlyIleTyrThrSerIleSerGluAlaLeuThrTrpIleGlnGluMet 311

RESULT 15
Q8CAN9 PRELIMINARY; PRT; 777 AA.
ID Q8CAN9
AC Q8CAN9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protease (Fragment).
GN PSS7 OR A130097D21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:54:56 ; Search time 22 seconds
(without alignments)
8227.084 Million cell updates/sec

Title: US-09-323-597C-1
Perfect score: 3179
Sequence: 1 ggcggagcgaggcgagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model p/US09323597/runat 06072004 111727 29642/app query.fasta 1.1927
-Q/cgn2 1/USPTO spoel p/US09323597/runat 06072004 111727 29642/app query.fasta 1.1927
-DB=SwissProt 42 -OPMT=fastan -SUFFIX=rsb -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597 @CNG 1.1 19 @runat 06072004 111727 29642 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2714	85.4	492	1 TMS2 HUMAN	Q15393 homo sapien
2	2117	66.6	490	1 TMS2 MOUSE	Q9jic8 mus musculus
3	880	27.7	453	1 TMS3 MOUSE	Q8k150 mus musculus
4	875.5	27.5	454	1 TMS3 HUMAN	P57727 homo sapien
5	680	21.4	435	1 TMS4 MOUSE	Q8vca5 mus musculus
6	676.5	21.3	437	1 TMS4 HUMAN	Q9nrs4 homo sapien
7	665.5	20.9	455	1 TMS5 MOUSE	Q9er04 mus musculus
8	660	20.8	1035	1 ENTK_BOVIN	P98072 bos taurus
9	651.5	20.5	457	1 TMS5 HUMAN	Q9h363 homo sapien
10	647	20.4	1019	1 ENTK_HUMAN	P98073 homo sapien
11	638	20.1	1069	1 ENTK_MOUSE	P97435 mus musculus
12	634	19.9	1034	1 ENTK_PIG	P98074 sus scrofa
13	601	18.9	1042	1 CORI HUMAN	Q9Y5Q5 homo sapien
14	579	18.2	1113	1 CORI_MOUSE	Q9Z319 mus musculus
15	575.5	18.1	855	1 ST14_HUMAN	Q9Y5Y6 homo sapien
16	569.5	17.9	638	1 KAL_HUMAN	P03952 homo sapien
17	568.5	17.9	855	1 ST14_MOUSE	P56677 mus musculus
18	567	17.8	417	1 HEPS_HUMAN	P05981 homo sapien

19	556	17.5	638	1 KAL_RAT	P14272 rattus norv
20	556	17.5	811	1 TMS6 MOUSE	Q9dbi0 mus musculus
21	552.5	17.4	812	1 PLAN_MOUSE	P20918 mus musculus
22	548.5	17.3	436	1 HEPS_MOUSE	C35453 mus musculus
23	538	16.9	638	1 KAL_MOUSE	P26262 mus musculus
24	536.5	16.9	811	1 TMS6 HUMAN	Q9iue0 homo sapien
25	533	16.8	625	1 FALL_HUMAN	P03951 homo sapien
26	531	16.7	416	1 HEPS_RAT	O05511 rattus norv
27	527	16.6	418	1 HATT_HUMAN	O60235 homo sapien
28	506	15.9	324	1 TEST_MOUSE	Q9jib7 mus musculus
29	501	15.8	421	1 ACRO_HUMAN	P10323 homo sapien
30	499.5	15.7	436	1 ACRO_MOUSE	P23578 mus musculus
31	496	15.6	422	1 DESI_HUMAN	Q9ul52 homo sapien
32	490.5	15.4	415	1 ACRO_PIG	P08001 sus scrofa
33	486.5	15.3	761	1 NCTR_MOUSE	O08762 mus musculus
34	486.5	15.3	790	1 PLAN_PIG	P06867 sus scrofa
35	486	15.3	810	1 PLAN_ERIEU	O29485 erinaceus e
36	483.5	15.2	786	1 STUB_DROME	Q05319 drosophila
37	479	15.1	311	1 TRYG_MOUSE	Q9qui7 mus musculus
38	475.5	15.0	431	1 ACRO_RABIT	P48038 oryctolagus
39	473.5	14.9	437	1 ACRO_RAT	P29233 rattus norv
40	473.5	14.9	810	1 PLAN_MACMU	P12545 macaca mula
41	471	14.8	290	1 PR27_HUMAN	Q9bq33 homo sapien
42	467.5	14.7	321	1 TRYG_HUMAN	Q9nrv2 homo sapien
43	465.5	14.6	306	1 BSS4_MOUSE	Q9er10 mus musculus
44	464.5	14.6	812	1 PLMN_BOVIN	P06868 bos taurus
45	463	14.6	273	1 TRYT_SHEEP	Q9xsm2 ovis aries

ALIGNMENTS

RESULT 1

TMS2_HUMAN
ID TMS2_HUMAN STANDARD; PRT; 492 AA.
AC O15393; Q9EXX1;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMRPS22 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM M.A.
RX MEDLINE=97468144; PubMed=9325052;
RA Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMRPS22 gene, which encodes a novel serine protease
with transmembrane, LDLR, and SRCR domains and maps to 21q22.3.";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309069; PubMed=1141763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=21139112; PubMed=11245484;
RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Safran D.C.,
RA Raitano A.B., Jakobovits A.;
RT "Catalytic cleavage of the androgen-regulated TMRPS22 protease results
in its secretion by prostate and prostate cancer epithelia.";
RL Cancer Res. 61:1686-1692(2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
RT "Expression of transmembrane serine protease TMRPS22 in mouse and
human tissues.";
RL J. Pathol. 193:134-140(2001).


```

QY 952 TGCGAGGCTCCATCATCACCCCGAGTGGATCGTGACGCGCCACTCGGTGGA AAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CTTCTTAACATCATCGGCTGAGCGGCTTTGCGGGGTTTGTGACAACTTTTCATG 1071
DB 301 ProLeuAsnAsnProThrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGATACCAAGTAGAAAAGTGTATCTCTCATCAATATGACTCCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProIleTyrAspSerLys 340
QY 1132 ACAAGAACATGATGCTGGCTGATGAAGCTGACAGAGCTCTGACTTTCAACGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACAGTGTCTGCGCCCAACCCAGGCGATGATGTCGACGACAGACGCTCTGCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGGCCACGAGGAGAAAGGGAAGACCTCAGAAGTGTGGAACGCTGCC 1311
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 ARGTCCTCTCTGATGACACACAGATGCAACAGCAGATGATCTATGACACCTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGACCATGATCTGTGCGGCTCTCTGCGAGGAACTGCAATCTTCCACAGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGAGGGCTCTGTGCTACTTGAAGACAAATATCTGTGCTGATGAGGGGATACAAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTCGCTGTGCAAGCTTACAGACACGAGGTGACGGAATGTGATGGTATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGTGATTTATCGACAAATGAGGCGCAGACGGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2
TMS2_MOUSE STANDARD; PRT; 490 AA.
AC Q9UJ08; Q9JRC4; Q9QY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE transmembrane protein X).
GN TMRPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c;
RX MEDLINE=20104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikio P.T.;
RT "Expression of transmembrane serine protease TMRPS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease x.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=BALB/c;
RX MEDLINE=20148617; PubMed=10683448;
RA Jacquinet E.J., Rao N.V., Rao G.V., Hoidal J.R.;
RT "Cloning, genomic organization, chromosomal assignment and expression
RT of a novel mosaic serine proteinase: epitheliasin.";
RL FEBS Lett. 468:93-100 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Mammary gland;
RX MEDLINE=22388457; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hoieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wuzey K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Willalson D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
CC cleavage and secreted (by similarity).
CC -1- TISSUE SPECIFICITY: Expressed mainly in prostate and kidney.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF199362; AAF97867.1; -
CC EXBL; AF243500; AAF64186.1; -
CC EXBL; AF113596; AAF21308.1; -
CC EXBL; BC038393; AAH38393.1; -
CC HSP; P00763; IDPO.
CC MEROPS; S01.247; -
CC MGD; MGI:1354381; Tmrps2.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; tryptsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR PROSITE; PS00420; SRCR_1; FALSE NEG.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253 TRANSMEMBRANE PROTEASE, SERINE 2, NON-

```

FT CHAIN 254 490 CATALYTIC CHAIN.
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CITOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT LDL-RECEPTOR CLASS A.
FT SRCR.
FT SERINE PROTEASE.
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CLEAVAGE (POTENTIAL).
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT INTERCHAIN (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT S -> L (IN REF. 3).
FT S -> N (IN REF. 3).
FT Y -> H (IN REF. 1).
FT N -> D (IN REF. 1).
SQ SEQUENCE 490 AA; 53479 NW; 07D2B03EA4D8A1A9 CRC64;

Alignment Scores:

Pred. No.: 2,648-144 Length: 490
Score: 2117.00 Matches: 385
Percent Similarity: 85.76% Conservative: 41
Best Local Similarity: 78.41% Mismatches: 63
Query Match: 66.59% Indels: 2
DB: 1 Gaps: 2

US-09-323-597C-1 (3-1738) x TMS2 MOUSE (1-490)

QY 112 ATGGCTTTGAATCAGGCTCACCAGCTATGGACCTTACTATGAAACCATGGATAC 171
DB 1 MetAlaLeuAenSer-GlySerProGlyIleGlyProCysTyrGluAsnHisGlyTyr 20
QY 172 CAACGGGAACCCCTATCCGCACAGCCACTGTGTCCTCCACCTGTCTACAGGTGAT 231
DB 21 GlnSerGluHisIleCysProArgProValAlaProAsnGlyTyrAsnLeuTyr 40
QY 232 CCGGCTCAGTACTACCTCCCGTCCCGTACGCGCCGAGGCTCTCAGCAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgIleThrGlnAla 60
QY 292 TCCAAACCCGCTGTCTGCAAGCCCAATCCCATCCCGGACAGTGTGACCTCAAG 351
DB 61 SerThrSerValIleHisThrHisProLysSer---SerGlyAlaProCysThrSerLys 79
QY 352 ACTAAGAAAGCACTGTGATCACCCTTGCCTTGGGACCTTCTCTGGGAGCTCGCGTG 411
DB 80 SerLysLysSerLeuCysLeuAlaLeuAlaLeuGlyThrValLeuThrGlyAlaAlaVal 99
QY 412 GCCTGCTGCTCTGCTGAACTTCAATGGGAGCAAGTGTCTCAACTCTGGATAGATGC 471
DB 100 AlaAlaValLeuLeuTyrArgPheTyrAspSerAsnCysSerThrSerGluMetGluCys 119
QY 472 GACTCTCAGTACTGATCAACCCCTCACTAAGTGTGTGATGGCTGTACACTGCCCC 531
DB 120 GlySerSerGlyThrCysIleSerSerSerLeuTyrCysAspGlyValAlaHisCysPro 139
QY 532 GCGGGGAGGAGCAAGATCGTGTGCTTCCCTCTACGGACCAAACTTCATCTTCAGGTG 591
DB 140 AsnGlyGluAspGluAsnArgCysValArgLeuTyrGlyGlnSerPheIleLeuGlnVal 159

RESULT 3

TMS3 MOUSE

ID TMS3 MOUSE STANDARD; PRT; 453 AA.

AC Q8K1T0; Q8VDB0;

QY 592 TACTCATCTCAGAGAGTCTCTGGGACCCCTGTGTGCCAGAGCTGGAACAGAACTAC 651
DB TyrSerGlnArgLysAlaIleTyrProValCysGlnAspAspTrpSerGluSerTyr 179
QY 652 GCGGGGCGGCTCAGGAGCATGGCTATAGGAATAATTTTACTCTTAGCAAGGAATA 711
DB GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLeu 199
QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACTGAACACAGTGGCGGCAATGTGAT 771
DB ProAspGlnSerGlyAlaThrSerPheMetLysLeuAsnValSerSerGlyAsnValAsp 219
QY 772 ATCTATAAAAACTGTATACCAAGTGTCTTCTTCAAAAGCAGTGGTTTCTTACGC 831
DB LeuTyrLysLysLeuTyrHisSerAspSerCysSerSerArgMetValValSerLeuArg 239
QY 832 TGTATAGCTCGCGGGTCAACTTGAACCTCAAGCCGACAGCAGAGTGTGGCGGCGAG 891
DB CysIleGluCysGlyValArg---SerValLysArgGlnSerArgIleValGlyLeu 258
QY 892 AGCGCGCTCCGCGGGCTGGCCCTGGCAGCTGACCTGCACGTCCAGAACGTCCAGTG 951
DB AsnAlaSerProGlyAspTrpProTrpGlnValSerLeuHisValGlnGlyValHisVal 278
QY 952 TGCAGAGCTCCATCATCACCCCGAGTGTGACAGCCGCCACTGCTGGGAAAAA 1011
DB CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluGlu 298
QY 1012 CCTCTTAACAATCCATGCGATTCGACGCGCTTGGCGGGATTTTGACAGCACTTTCATG 1071
DB ProLeuSerGlyProArgTyrTrpThrAlaPheIleGlyIleLeuArgGlnSerLeuMet 318
QY 1072 TTCTATGAGCCGGATACCAAGTAGAAAAAGTGTATTTCTCATCCAAATTTATGATCCAAG 1131
DB PheTyrGlySerArgHisGlnValGluLysValIleSerHisProAsnTyrAspSerLys 338
QY 1132 ACCAAGAACATGACATTCGCTGATGAGCTGACAGCTCAGAGGCTCTGCTTCAAGACCTA 1191
DB ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnThrProLeuAlaPheAsnAspLeu 358
QY 1192 GTGAACACAGTGTCTGCCCAACCCAGGATGATGTGCGCCAGACAGCTCTGCTGG 1251
DB ValLysProValCysLeuProAsnProGlyMetMetLeuAspLeuAspGlnGluCysTrp 378
QY 1252 ATTTCGGTGGGGGCGCACCGAGGAGAAAGGAGACCTCAGAGTGTCTGAGCGTGGC 1311
DB IleSerGlyTrpGlyAlaThrTyrGluLysGlyLysThrSerAspValLeuAsnAlaAla 398
QY 1312 AAGGTGCTTCTCATTCAGACACAGAGATGCAACAGACAGATATGTCTATGACCACTGATC 1371
DB MetValProLeuIleGluProSerLysCysAsnSerLysTyrIleTyrAsnAsnLeuIle 418
QY 1372 ACACAGCCATGATGTGTGCGGCTTCTGCGGGGAACTCGATTCCTCCAGGGTGAC 1431
DB ThrProAlaMetIleCysAlaGlyPheLeuGlnGlySerValAspSerCysGlnGlyAsp 438
QY 1432 AGTGGAGGGCTCTGGTCACTTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAAAG 1491
DB SerGlyGlyProLeuValThrLeuLysAsnGlyIleTrpTrpLeuIleGlyAspThrSer 458
QY 1492 TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATC 1551
DB TrpGlySerGlyCysAlaLysAlaLeuArgProGlyValTyrGlyAsnValThrValPhe 478
QY 1552 ACCGACTGGATTTTACGACAAATGAGGCGACAC 1584
DB ThrAspTrpIleTyrGlnGlnMetArgAlaAsn 489

QY 1000 TCGCTGCGAAACCTCTTACCAATCGATCGCATTCGGACGGCATTTGCGGGATTTTGAGA 1059
Db CysVal---TyrAspLeuTyrHisProLysSerTrpThrValGlnValGlyLeuVal--- 275
QY 1060 CAATCTTTCATCTTCTATGAGGCC---GGATACCAAGTAGAAAAGTGAATTTCTCATCCA 1116
Db ---SerLeuMetAspSerProValProSerHisLeuValGluLysIleIleTyrHisSer 294
QY 1117 AATTATGACTCCAGACCGAGCAATGATCGCTGATGATGATGATGATGATGATGATGATG 1176
Db ---LysTyrLysProLysArgLeuGlyAsnAspIleAlaLeuMetLysLeuSerGluProLeu 314
QY 1177 ACTTTCACGACCTAGTGAACACCTGCTGCTGCCAACCCAGCCAGGCATGCTGCGAGCCA 1236
Db ---ThrPheAspGluThrIleGlnProLysCysLeuProAsnSerGluGluAsnPheProAsp 334
QY 1237 GAACAGCTCTGCTGATTTCCGGTGGGGCCACCGAGGAGAAAGGAGACCTTCAGAA 1296
Db ---GlyLysLeuCysTrpThrSerGlyTrpGlyAlaThrGluAspGlyGlyAspAlaSerPro 354
QY 1297 GTGCTGAACGCTGCCAAGGTCTTCTCATTCGACACAGACAGATGCAACAGCAGATATGTC 1356
Db ---ValLeuAsnHisAlaAlaValProLeuIleSerLeuLysIleCysAsnHisArgAspVal 374
QY 1357 TATGCAACCTGATCACACCGACCATGATCTGCGCGCTTCTGCGAGGGAAGCTCGAT 1416
Db ---TyrGlyGlyIleIleSerProSerMetLeuCysAlaGlyTyrLeuLysGlyGlyValAsp 394
QY 1417 TCTTCCAGGCTGACAGTGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
Db ---SerCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluArgLeuGlyLeuLysLeu 414
QY 1477 ATAGGGGATACAGCTGGGGTTCTGCTGCTGCCAAGCTTACAGACGACGAGGTGACGGG 1536
Db ValGlyAlaThrSerPheGlyIleGlyCysAlaGluValAsnLysProGlyValTyrThr 434
QY 1537 AATGTGATGATATCACGACCTGATTTATGCAATATGAGGGGAGAC 1584
Db ---ArgIleThrSerPheLeuAspTrpIleHisGluGlnLeuGluArgAsp 450

RESULT 4
TMS3 HUMAN
ID TMS3 HUMAN STANDARD; PRT; 454 AA.
AC P5727;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
TAG-12) (Tumor associated differentially-expressed gene-12 protein).
GN TMPSR3 OR TAG12 OR ECHS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS A AND T).
RP TISSUE=Ovarian carcinoma;
RX MEDLINE=20521358; PubMed=11068177;
RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider B.N.,
RA Wang Y., Farley T.H., O'Brien T.J.;
RT "Ovarian tumor cells express a novel multi-domain cell surface serine
protease.";
RL Biochim. Biophys. Acta 1502:337-350(2000).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS A, B AND D), AND VARIANT ILE-53.
RX MEDLINE=20578749; PubMed=11137999;
RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrost R.,
RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
RA Younus F., Mehdi S.O., Radhakrishna U., Papasavas M.P., Gehrig C.,
RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
RA Antonarakis S.E.;
RT "Insertion of beta-satellite repeats identifies a transmembrane
protease causing both congenital and childhood onset autosomal

recessive deafness.";
Nat. Genet. 27:59-63(2001).
[3]
RP SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
RX MEDLINE=22281255; PubMed=12393794;
RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Doughterty L., Vuagniaux G., Guidi E., Okui M., Rossier C., Hancock M.,
RA Buchet K., Raymond A., Hummel E., Marzella P.L., Kudoh J.,
RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
RT "The transmembrane serine protease (TMPSR3) mutated in deafness
DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro.";
Hum. Mol. Genet. 11:2829-2836(2002).
[4]
RP VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.
RX MEDLINE=21354487; PubMed=11462234;
RA Masoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Grattini M.,
RA Papasavas M.P., Driha M., Eigaied-Bouilla A., Wattenhofer M.,
RA Rossier C., Scott H.S., Ayadi H., Guipponi M.;
RT "Novel missense mutations of TMPSR3 in two consanguineous Tunisian
families with non-syndromic autosomal recessive deafness.";
Hum. Mutat. 18:101-108(2001).
[5]
RP VARIANTS DFNB8/DFNB10 TRP-109, PHE-194 AND ARG-407, AND VARIANTS
ILE-53; SER-111 AND VAL-253.
RX MEDLINE=21317610; PubMed=11424922;
RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
RA Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B.,
RA Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
RA Nance W.S., Wilcox R.R., Friedman T.B., Morell R.J.;
RT "Novel mutations of TMPSR3 in four DFNB8/B10 families segregating
congenital autosomal recessive deafness.";
J. Med. Genet. 38:396-400(2001).
[6]
RP VARIANTS DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
RX MEDLINE=21904597; PubMed=11907649;
RA Wattenhofer M., Di Iorio V., Rabinet R., Dougherty L., Panpanos A.,
RA Schwede T., Montserrat-Sentis B., Arbones L., Iliades T.,
RA Pasquabiscaglia A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,
RA Petersen M.B., Estivill X., Gasparini P., Scott H.S.,
RA Antonarakis S.E.;
RT "Mutations in the TMPSR3 gene are a rare cause of childhood
nonsyndromic deafness in Caucasian patients.";
J. Mol. Med. 80:124-131(2002).
RL J. Med. Genet. 38:396-400(2001).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
ENaC.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
reticulum.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=P57727-1; Sequence=Displayed;
CC Name=B; Synonyms=C;
CC IsoId=P57727-2; Sequence=VSP_005391;
CC Name=D;
CC IsoId=P57727-3; Sequence=VSP_005392;
CC Name=T; Synonyms=Truncated; TAG-12V;
CC IsoId=P57727-4; Sequence=VSP_005393; VSP_005394;
CC -!- TISSUE SPECIFICITY: Expressed in many tissues including fetal
cochlea. Isoform T is found at increased levels in some
carcinomas.
CC -!- PTM: Undergoes autolytic activation.
CC -!- DISEASE: Defects in TMPSR3 are a cause of childhood-onset
autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
CC -!- DISEASE: Defects in TMPSR3 are a cause of congenital autosomal
recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

QY 955 GGAGGCTCATCATCAGCCCGGAGTGGATGCTGAGCAGCGGCCCATCTGGTGGAAAAACCT 1014
 Db 243 GGGGlySerValIleThrProLeuTrpIleIleThrAlaAlaHisCysVal---TyrAsp 261
 QY 1015 CTTAAACAATCCATGCGCATTTGGACGCGCAATTTGGACGCAATCTTTCATGTTTC 1074
 Db 262 LeuTyrLeuProLeuSerTrpThrIleGlnValGlyLeuVal-----SerLeuLeuAsp 279
 QY 1075 TATGGAGCC---GGATACCAAGTAGAAAGTATTTCTCATCCAAATAGTACTCCCAAG 1131
 Db 280 AsnProAlaProSerHisLeuValGluValGlyValTyrHisSerLysTyrLysProLys 299
 QY 1132 ACCAAGAACATGATGCGTGTATGATGAGTGCAGAGCTCTCTGACTTTCACGACCTA 1191
 Db 300 ArgLeuGlyAsnAspIleAlaLeuMetLysLeuAlaGlyProLeuIleThrPheAsnGluMet 319
 QY 1192 GTGAACACAGTGTCTCCCAACCCAGGCGATGATGTCGACGCGACAGCTCTGCTGG 1251
 Db 320 IleGlnProValCysLeuProAsnSerGluGluAsnPheProAspGlyLysValCysTrp 339
 QY 1252 ATTTCCGGGTGGGGCCACCGAGGAG---AAAGGAGAGCTCAGAGTCTGCAAGCT 1308
 Db 340 ThrSerGlyTrpGlyAlaThrLeuAspGlyAlaGlyAspAlaSerProValLeuAsnHis 359
 QY 1309 GCCAAGTGTCTCTCATGACACAGATGCAACGATGATGCTATGACAACTG 1368
 Db 360 AlaAlaValProLeuIleSerAsnLysIleCysAsnHisAAspValTyrGlyIle 379
 QY 1369 ATCACACAGCATGATGTCGGCTCTCTGACGAGGACGTCGATTTGTCAGGT 1428
 Db 380 IleSerProSerMetLeuCysAlaGlyTyrLeuThrGlyGlyValAspSerCysGlnGly 399
 QY 1429 CACAGTGGAGGCGCTCTGCTCACTTCGAAGAACATATCTGCTGGCTAGAGGGATACA 1488
 Db 400 AspSerGlyGlyProLeuValCysGlnGluArgIleTyrLysLeuValGlyAlaThr 419
 QY 1489 ACCTGGGGTCTGCTGTCGCAAGCTTACAGACAGAGTGTACGGGAATGTATGATGTA 1548
 Db 420 SerPheGlyIleGlyCysAlaGluValAsnLysProGlyValTyrThrArgValThrSer 439
 QY 1549 TTCACGGAAGTGGATTATTCGAACAATGAGGCGACAG 1584
 Db 440 PheLeuAspTrpIleHisGluGlnMetGluArgAsp 451
 RESULT 5
 TMS4 MOUSE
 ID TMS4 MOUSE
 AC Q8VCA5
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Channel-activating
 DE protease 2) (mCAP2).
 GN TMPSR4 OR CAP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22144321; PubMed=12149280;
 RA Vagniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
 RT "Synergistic activation of ENaC by three membrane-bound channel-
 RT activating serine proteases (mCAP1, mCAP2, and mCAP3) and serum- and
 RT glucocorticoid-regulated kinase (SGK1) in Xenopus oocytes.";
 RL J. Gen. Physiol. 120:191-201(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Breast tumor;
 RC MEDLINE=22389257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gamaralane P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez R.C., Touchman J.W., Green E.D., Dickson M.C.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.B.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENaC.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AY043240; AAK85307.1; -;
 DR EMBL; BC021368; AAK21368.1; -;
 DR HSSP; P00761; IAN1.
 DR MEROPS; S01.034; -;
 DR MGD; MGI:2384877; Tmprss4.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1-
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS01209; LDLRA_1; FALSE NEG.
 DR PROSITE; PS00668; LDLRA_2; FALSE NEG.
 DR PROSITE; PS00420; SRCR_1; FALSE NEG.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin HIS; 1.
 DR PROSITE; PS00135; TRYPsin SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 30 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 31 51 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 52 435 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 59 101 LDL-RECEPTOR CLASS A.
 FT DOMAIN 102 202 SRCR.
 FT DOMAIN 203 435 SERINE PROTEASE.
 FT ACT_SITE 243 243 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 288 288 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 385 385 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 202 203 CLEAVAGE (POTENTIAL).
 FT DISULFID 62 81 BY SIMILARITY.
 FT DISULFID 75 90 BY SIMILARITY.
 FT DISULFID 125 181 BY SIMILARITY.
 FT DISULFID 138 191 BY SIMILARITY.
 FT DISULFID 194 308 BY SIMILARITY.
 FT DISULFID 228 244 BY SIMILARITY.
 FT DISULFID 354 370 BY SIMILARITY.
 FT DISULFID 381 408 BY SIMILARITY.

FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 435 AA; 47495 MW; DC52B45A43E01369 CRC64;
Alignment Scores:
Pred. No.: 3.21e-41 Length: 435
Score: 680.00 Matches: 163
Percent Similarity: 50.56% Conservative: 61
Best Local Similarity: 36.79% Mismatches: 173
Query Match: 21.39% Indels: 46
DB: 1 Gaps: 13
US-09-323-597C-1 (1-1738) x TMS4_MOUSE (1-435)
QY 313 CAGCCCAATCCCATCGGACAGTGTGCACCTCAAGACTAGAAGAAGCA----- 363
Db LysProArgArgProGlnGluThrPhe-----LysLysValGlyIlePro 33
QY 364 ---CTGGCATCACTACCTCGGACCTTCCCTCGTGGGAGTCTCGGCGCGCTGGC 420
Db IleIleAlaValLeuLeuSerLeuIleAlaValIleValAlaLeuLeuIleLysVal 53
QY 421 CTACTCTGGAGTTCATCGGACAGCAAGTGTCTCAACTCTGGGATAGAGTGGACTCTCA 480
Db IleLeuAspLysIleTyr-----PheIleCysGlySerPro 65
QY 481 GGTACTGTCATCAACCCCTCTAATCTGGTGTGATGGGTGTGCACACTGCCCGGGGGAG 540
Db LeuThrPheIleGlnArgGlyGlnLeuCysAspGlyIleLeuAspCysAlaSerGlyGlu 85
QY 541 GACGAGAAATCGGTGT-----GTGCGCTCTAC 567
Db AspGluGluHisCysValLysAspPheProGluLysProGlyValAlaValArgLeuSer 105
QY 568 GGACCAAACTCATCTCTAGTGTACTCATCTCAGAGGAGTCTCGCACCTGTGTC 627
Db LysAspArgSerThrLeuGlnValLeuAspAlaAlaThrGlyThrTrpAlaSerValCys 125
QY 628 CAAGACGACTGGAACGAGAACTACGGCGGGCGCTCGACGGACATCGGCTATAGAAT 687
Db PheAspAsnPheThrGluAlaLeuAlaLysThrAlaCysArgGlnMetGlyTyrAspSer 145
QY 688 AAT---TTTACTCTAGCAGGAAATAGTGTGATCAGCAGGATCCACCGCTTATGAAA 744
Db GlnProAlaPheArgAlaValGluIleArgProAspGlnAsn-----LeuPro 161
QY 745 CTGAACACAACTGCGCGCAATGTCGATATCTATAAAACCTGACACAGTGAGCTGT 804
Db ValAlaGlnValThrGlyAsnSerGlnGluLeuGlnValGlnAsnGlySerArgSerCys 181
QY 805 TCTTCAAAAGCAGTGTCTTTTACGTGTATAGCTGCGGGGTCACTTGAATCAAGC 864
Db LeuSerGlySerLeuValSerLeuArgCysLeuAspCysGlyLysSerLeu----- 198
QY 865 CGCCAGACGAGATTGTGGGGCGGAGAGCGCGCTCCCGGGCGCTCGCTGCGAGTGC 924
Db LysThrProArgValValGlyValGluAlaProValAspSerTrpProGlnVal 218
QY 925 AGCTGCACTGCCAGAGCTCCAGTGTGGAGGCTCCATCATCAACCCCGAGTGGATC 984
Db SerIleGlnTyrAsnLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpIle 238
QY 985 GTGACAGCGCCCTCGTGGAAACCTTAAAC---AATCATGGCATTCGACCGCA 1041
Db LeuThrAlaAlaHisCysPheArgLysTyrLeuAspValSerTrpLysValArgAla 258
QY 1042 TTTCGGGGGATTTTGAGCAATCTTTCATGTCTATGAGCGCGGATACCAAGTAGAAAA 1101
Db GlySerAsnIleLeuGlyAsnSer-----ProSerLeuProValAlaLys 273
QY 1102 GTG---ATTCTCATCCAAATATGATCTCCAGACCAAGCAATGATGCGCTGATG 1158
Db IlePheIleAlaGluProAsn---ProLeuTyrProLysGluLysAspIleAlaLeuVal 292

RESULT 6

TMS4_HUMAN
ID TMS4_HUMAN STANDARD; PRT; 437 AA.
AC Q9NR84; Q9NZAS;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
DE protease 2) (WP-SP2).
GN TMPSR4 OR TMPSR3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenburger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMPSR3) overexpressed in
RT pancreatic cancer";
RL Cancer Res. 60:2602-2606(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Sneekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
RT trachea, colon, and small intestine: identification, cloning, and
RT chromosomal localization";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,


```
QY 487 TGCATCAACCCCTCTAACTGGTGTGATGGCGTGTACACTGCGCGCGGAGACGAG 546
Db 84 LeuGlnGluGlu- - - - -GluMetThrLeuAsnCySerProGlyValSerCysGlu 99
QY 547 AATCGGTGTGTT- - - - -CGCCTCTACGACCAACATTC 579
Db 100 GluGluLeuLeuProSerLeuProLysThrValSerPheArgIleAsnGlyGluAspLeu 119
QY 580 ATCCCTCAGGTACTCATCTCAGAGGAGTCTGCGACCGTGTGTGCGACAGACACTGG 639
Db 120 LeuLeuGlnValGlnValArgAlaArgProAspTrpLeuLeuValCysHisGluGlyTrp 139
QY 640 AACGAGAACTACGGCGCGCGCTGCAGGACATGGGCTAT- - -AAGAATAATTTTAC 696
Db 140 SerProAlaLeuGlyMethIleCysLysSerLeuGlyHisIleArgLeuThrGlnHis 159
QY 697 TCTAGCCAAAGGATAGTGATGAC- - - - -ACCGATCCACCGCTTTATGAACCTGAAC 750
Db 160 LysAlaValAsnLeuSerAspIleLysLeuAsnArgSerGlnGluPheAlaGlnLeuSer 179
QY 751 ACAAGTCCGCGCAATGTGCTATATCTATAAAACATGTACACAGCATGCTGTTCTTCA 810
Db 180 AlaArgProGlyGlyLeu- - -ValGluGluAlaTrpLysProSerAlaAsnCySerPro 198
QY 811 AAAGCAGTGGTCTTCTTACGTGTATAGCTGCGGGTCAACTGGAATCAAGCGCCGAC 870
Db 199 GlyArgIleValSerLeuLysCysSerGlyCysGlyAlaArg- - - - -ProLeuAla 215
QY 871 ACAGCATTTGCGGGCGGAGCGCGCTCCCGGGGCTGCGCTGCGAGCTCAGCTG 930
Db 216 SerArgIleValGlyGlnAlaValAlaSerGlyArgTrpProTrpGlnAlaSerVal 235
QY 931 CACGTCAGAGCTCCATCGTGTGCGAGGCTCCATCATCACCCCGAGTGGATCGTACA 990
Db 236 MetLeuGlySerArgHisThrCysGlyAlaSerValLeuAlaProHisTrpValValThr 255
QY 991 GCGGCGCATCGGTGTGAAAA- - -CCTCTTACAATCCATGCGCATTCGACGCGCATTCGG 1047
Db 256 AlaAlaHisCysMetTyrSerPheArgLeuSerArgLeuSerSerTrpArgValHisAla 275
QY 1048 GGGATTTTGAGACAATCTTTTCATGTCTTATGAGCGCGGATACCAAGTAAAGATGATT 1107
Db 276 GlyLeuValSerHisGlyAlaValArgGlnHisGlnGlyThrMetValGluLysIle 295
QY 1108 TCTCATCAATATGACTCCAGACCAAGACATGATGCTGCTGCTGATGAGCTGAG 1167
Db 296 ProHisProLeuTyrSerAlaGlnAsnHisAspTyrAspValAlaLeuLeuGlnLeuArg 315
QY 1168 AAGCCTCTGACTTTCAACGACCTAGTGAACACAGTGTGTCTGCGCAACCCAGCATGATG 1227
Db 316 ThrProIleAsnPheSerAspThrValAspAlaValCysLeuProAlaLysGluGlnTyr 335
QY 1228 CTGCGACCCAGAACAGCTCTGCTGGATTCGGGTGCGGGGCGCCACCGAG- - -GAGAAAGG 1284
Db 336 PheProTrpGlySerGlnCysTrpValSerGlyTrpGlyHisThrAspProSerHisThr 355
QY 1285 AAGACTCAGAAGTGTGAACGCTCCAGGTGCTTCTCATTTGACACAGAGATGCAAC 1344
Db 356 HisSerSerAspThrLeuGlnAspThrMetValProLeuLeuSerThrHisLeuCysAsn 375
QY 1345 ACCGATATGCTATGACACCTGATCACCAGCCATGATCTGTGCGCGTCTCTGCGAG 1404
Db 376 SerSerCysMetTyrSerGlyAlaLeuThrHisArgMetLeuCysAlaGlyTyrLeuAsp 395
QY 1405 GGGACGCTGATCTTCCAGGTCACAGTGCAGGCGCTGCTCACTTCGAGAACAT 1464
Db 396 GlyArgAlaAspAlaCysGlnGlyAspSerGlyGlyProLeuValCysProSerGlyAsp 415
QY 1465 ATCTGGTGGCTGATAGGGATACAAAGCTGGGTTCTGCTGTGCCAAAGCTTACAGACA 1524
Db 416 ThrTrpHisLeuValGlyValValSerTrpGlyArgGlyCysAlaGluProAsnArgPro 435
```

```
QY 1525 GGACTGTACGGAAATGTGATGGTATTTCACGCACTGGATTAT 1566
Db 436 GlyValTyrAlaLysValAlaGluPheLeuAspTrpIleHis 449
RESULT 8
ENTK_BOVIN
ID ENTK_BOVIN STANDARD; PRT: 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu O., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;
RA Lavallie B.R., Rehmulla A., Racie L.A., Dibiaseo E.A.,
RA Perenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -!- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multichain (heavy) chain linked by a disulfide bond.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P98072-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P98072-2; Sequence=VSP_005386;
CC -!- TISSUE SPECIFICITY: Intestinal brush border.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPsin-LIKE PROTEASE.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SSCR domain.
CC -!- SIMILARITY: Contains 1 SSCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```


Db 772 GlnCysLeuGluAspSerLeuLeuLeuGlnCysAsnTyrLysSerCysGlyLysLys 791
 QY 853 TTGAACCTAAGCCGACAGC---AGGATTCTGGCGGCGAGAGCGCGCTCCCGGGCC 909
 Db 792 LeuValThrGlnGluValSerProLysIleValGlyGlySerAspSerArgGluGlyAla 811
 QY 910 TGGCCCTGGAGGTGAGCTCCAGCTCCAGAACCTCCACGCTGCGGAGGCTCCATCATC 969
 Db 812 TrpProTrpValValAlaLeuTyrPheAspAspGlnGlnValCysGlyAlaSerLeuVal 831
 QY 970 ACCCCGAGTGGATGTCACAGCGCCCACTGCTGGGAAACCTCTTAACATCCATGG 1029
 Db 832 SerArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnMetGluProSer 851
 QY 1030 CATTCGAGCATTTGGCGGGATTTTCAGACAACTTTTCATGCTCTTCATGGCCGGATAC 1089
 Db 852 LysTrpLysAlaValLeuGlyLeuHisMetAlaSerAsnLeu-----ThrSerPro 868
 QY 1090 CAAGTAGA-----AAAGTGATTCTCATCCAAATTTATGACTCCAGACC 1134
 Db 869 GlnIleGluThrArgLeuLeuAspGlnIleValIleAsnProHisTyrAsnLysArgArg 888
 QY 1135 AAGAACATGATGCTGCTGATGAGCTGAGCTGAGAGGCTCTGACTTTCAACGACCTAGTG 1194
 Db 889 LysAsnAsnAspIleAlaMetMetHisLeuGluMetLysValAsnTyrThrAspTyrIle 908
 QY 1195 AAACAGTGTCTGCTCCCAACCCAGGAGATGCTGCGAGCGAACAGCTGCTGCTGAT 1254
 Db 909 GlnProIleCysLeuProGluGluAsnGlnValPheProGlyArgIleCysSerIle 928
 QY 1255 TCCGGTGGGGGGCCACCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1314
 Db 929 AlaGlyTrpGlyAlaLeuIleTyrGlnIleSerThrAlaAspValLeuGlnGluAlaAsp 948
 QY 1315 GTGCTTCTCATGACAGACAGATGCAACAGACAGATGCTGATGATGCAACCTGATCACA 1374
 Db 949 ValProLeuLeuSerAsnGluLysCysGlnGlnGlnMetProGluTyrAsn---IleThr 967
 QY 1375 CCAGCCATGACTGTCGGGCTTCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434
 Db 968 GluAsnMetValCysAlaGlyTyrGluAlaGlyValAspSerCysGlnGlyAspSer 987
 QY 1435 GGAGGCTCTGCTGCTCGAAGAACATATCTGGTGGCTGATAGGGGATACAGCTGG 1494
 Db 988 GlyGlyProLeuMetCysGlnGluAsnAsnArgTrpLeuLeuAlaGlyValThrSerPhe 1007
 QY 1495 GGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1554
 Db 1008 GlyTyrGlnCysAlaLeuProAsnArgProGlyValTyrAlaArgValProArgPheThr 1027
 QY 1555 GACTGGATT 1563
 Db 1028 GluTrpIle 1030

RESULT 9

TMS5_HUMAN STANDARD; PRT; 457 AA.
 ID Q9H3S3;
 AC Q9H3S3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMS5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=11741986;
 RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
 RT "Spinesin/TMS5, a novel transmembrane serine protease, cloned from

RT human spinal cord.";
 RL J. Biol. Chem. 277:6806-6812(2002).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -1- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
 CC neurons, in their axons, and at the synapses of motoneurons in the
 CC spinal cord.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AB028140; BAB20375.1; -;
 DR RSP; P00763; IDPO.
 DR Genew; HGNC:14908; TMS5.
 DR MIM; 606751; -;
 DR MEROPS; S01.313; -;
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM0020; Tryp_SPC; 1.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin HIS; 1.
 DR PROSITE; PS00135; TRYPsin SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE NEG.
 DR PROSITE; PS0287; SRCR_2; FALSE NEG.
 DR HydroLase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein.
 FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 71 457 (POTENTIAL).
 FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 218 457 SRCR.
 FT ACT_SITE 258 258 SERINE PROTEASE.
 FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 217 218 CLEAVAGE (POTENTIAL).
 FT DISULFID 135 196 BY SIMILARITY.
 FT DISULFID 148 206 BY SIMILARITY.
 FT DISULFID 209 328 BY SIMILARITY.
 FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;
 Alignment Scores:
 Pred. No.: 3,59e-39 Length: 457
 Score: 651.50 Matches: 157
 Percent Similarity: 46.65% Conservative: 59
 Best Local Similarity: 33.91% Mismatches: 203
 Query Match: 20.49% Indels: 44
 DB: 1 Gaps: 10

US-09-323-597C-1 (1-1738) x TMS5_HUMAN (1-457)

QY 246 CCGTCCCGGTCGCCAGTACGCCCGAGGGTCTGACGAGCTTCCAAACCCGCTCGT 305

Db 9 ProProMetGluAlaGlnTyrAlaGluGlyProGlyProGlyIlePheArgAlaGlu 28

QY 306 CTGACGCGCCCAATCCCATCCCGGACAGTGTGACCTCAAGACTAAGAAGCACT 365
 Db 29 ProGlyAspGlnHisProIleSerGlnAlaVal-CysTrpArgSerMetArgGgl 48
 QY 366 GTGCATCACCTTGACCTCGGGGACCTCTCTCGTGGAGCTCGGCTGGCGTGCCTACT 425
 Db 48 yCysAlaValLeuGlyAlaLeuGlyLeuAlaGlyAlaGlyValGlySerTrpLeuLe 68
 QY 426 CTGAAAGTTCATGGCGCAGCAAGTGTCTCAAC-----TCTGGATAGAGTCGCTCCTC 479
 Db 68 uValLeuTyrLeuCysProAlaAlaSerGlnProIleSerGlyThrLeuGlnAspGluGl 88
 QY 480 AGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGTGCACACTGCCCGCGCGGGA 539
 Db 88 uileThrLeu-----SerCysSerGluAlaLe 97
 QY 540 GGACGAGATCGGTGTGT-----CGCTCTACGGGACC 572
 Db 97 rAlaGluGluAlaLeuLeuProAlaLeuProLysThrValSerPheArgIleAsnSerGl 117
 QY 573 AAATCTCATCTCTCAGGTGTACTCATCTCAGAGGAAGTCTCGGCACCTGTGTGCCAAGA 632
 Db 117 uAspPheLeuLeuGluAlaGlnValArgAspGlnProArgTrpLeuValCysHisGl 137
 QY 633 CGACTGAACAGAGAACTACGGCGCGCGCTCGAGGACATCGGCTAT---AGAAATAA 689
 Db 137 uGlyTrpSerProAlaLeuGlyLeuGlnIleCysTrpSerLeuGlyHisLeuArgLeuTh 157
 QY 690 TTTTACTCTAGCAAGAGATAGTGGATGACAGCGGATCCACAGCTTTATGAACAGAA 749
 Db 157 rHisHisLysGlyValAsnLeuThrAsp-----lleLysLeuAs 170
 QY 750 CACAAGTCCCGCAATCTCATATCTATAAAAAACTG-----TACCACAGTGTATGCC-- 801
 Db 170 nSerSerGlnGluPheAlaGlnLeuSerProArgLeuGlyGlyPheLeuGluAlaTr 190
 QY 802 -----TGTTCTTCAAAGCAGTGGTTTCTTTTACCTGTGTATAGCTCGCG 845
 Db 190 pGlnProArgAsnAsnCysThrSerGlyGlnValSerLeuArgCysSerGluCysGl 210
 QY 846 GGTCAACTTGAACTCAAGCCCGCAGAGATGTGGCGGCGAGAGCGGCTCCCGGG 905
 Db 210 yAlaArg-----ProLeuAlaSerArgIleValGlyGlyGlnSerValAlaProGl 227
 QY 906 GGCTGCTCCCTGGAGTGCAGCTGCAGCTCCAGAACTCCAGTGTGGAGGCTCCAT 965
 Db 227 yArgTrpProIleGlnAlaSerValAlaLeuGlyPheArgHisThrCysGlyGlySerVa 247
 QY 966 CATCACCCCGAGTGGATGTGACAGCGGCCACTGCTGGAAAAA---CCTCTTAACAA 1022
 Db 247 lLeuAlaProArgTrpValThrAlaAlaHisCysMetHisSerPheArgLeuAlaAr 267
 QY 1023 TCCATGGCATTTGGCGGCATTTGGGGGATTTTGGAGACAATCTTTCATGTCTATGGAGC 1082
 Db 267 gLeuSerSerTrpArgValHisAlaGlyLeuValSerHisSerAlaValArgProHisGl 287
 QY 1083 CGGATACCAAGTAGAAAAATGATTTCTCTCCAAATATGATCTCCAGACCCAGAACAA 1142
 Db 287 nGlyAlaLeuValGluArgIleIleProHisProLeuTyrSerAlaGlnAsnHisAspTy 307
 QY 1143 TGCAATTCGCTGATGAAGTGCAGAGCGCTCTGACCTTCAACACACTAGTGAACCCAGT 1202
 Db 307 rAspValAlaLeuLeuArgLeuGlnThrAlaLeuAsnPheSerAspThrValGlyAlaVa 327
 QY 1203 GTGTCTGCCCAACCCAGCGCATGTGTCAGCGAGCAACAGCTCTGTCTGGATTTCCGGGTG 1262
 Db 327 lCysLeuProAlaLysGluGlnHisPheProLysGlySerArgCysTrpValSerGlyTr 347
 QY 1263 GGGGGCCACCGAGGAGGAAGGAG---ACCTCAGAAGTGTCTCAACGCTCCCAAGTGTCT 1319
 Db 347 pGlyHisThrHisProSerHisThrTyrSerSerAspMetLeuGlnAspThrValValPr 367
 QY 1320 TCTCATTGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACACAGC 1379

Db 367 oLeuPheSerThrGlnLeuCysAsnSerSerCysValTyrSerGlyAlaLeuThrProAr 387
 QY 1380 CATGATCTGTGGCGGCTCTCTCGAGGACAGCTCGATTCTTCGAGGTCACAGTGGAGG 1439
 Db 387 gMetLeuCysAlaGlyTyrLeuAspGlyArgAlaAspAlaCysGlnGlyAspSerGlyGl 407
 QY 1440 GCCTCTGTCTCCTTCGAGAACCAATATCTGTGGCTGTATGGGATACAAAGCTGGGTTTC 1499
 Db 407 yProLeuValCysProAspGlyAspThrTrpArgLeuValGlyValValSerTrpGlyAr 427
 QY 1500 TGCGTGTGCCAAGCTTACAGACACAGAGTGTACGGGAATGTGATGTATTCACGAGCTG 1559
 Db 427 gAlaCysAlaGluProAsnHisProGlyValThrAlaLysValAlaGluPheLeuAspTr 447
 QY 1560 GATTTAT 1566
 Db 447 pIleHis 449
 RESULT 10
 ID ENTX_HUMAN STANDARD; PRT; 1019 AA.
 AC P98073;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTX
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Duoenum;
 RX MEDLINE=95234679; PubMed=7718557;
 RA Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 RT the proteolytic activator of trypsinogen.";
 RL Biochemistry 34:4562-4568(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND DISEASE.
 RX MEDLINE=21606074; PubMed=11719902;
 RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
 RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
 RT "Mutations in the proenteropeptidase gene are the molecular cause of
 RT congenital enteropeptidase deficiency.";
 RL Am. J. Hum. Genet. 70:20-25(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Dalabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Kennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Leinhardt H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=Duoenum;
 RX MEDLINE=94329561; PubMed=8052624;
 RA Kitamoto Y., Iuan X., Wu Q., McCourt D.W., Sadler J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 RT protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).

PT	CONFLICT	754	771	SEQCQDSLRPLQCNHKS -> RRNAXNEIDALSPITLLIA (IN REF. 3).
PT				
Alignment Scores:				
Pred. No.:	8.93e-39	Length:	1019	
Score:	647.00	Matches:	138	
Percent Similarity:	53.25%	Conservative:	67	
Best Local Similarity:	35.84%	Mismatches:	154	
Query Match:	20.35%	Indels:	26	
DB:	1	Gaps:	10	
US-09-323-597C-1 (1-1738) x ENT_K_HUMAN (1-1019)				
Qy	448	TGCTCCAACTCTGGGATAGACTGGAGCTCTCTCAGCTACTCTGCATCAACAACCCCTCTAACTGG	507	
Db	643	CysLysAlaAspHisPheGlnCys---LysAsnGlyGluCysValProLeuValAsnLeu	661	
Qy	508	TGTGATGGCGTGTCACTGCTCCCGGGGAGGACGAGAAATCGGTGTTCGCTCTTAC	567	
Db	662	CysAspGlyHisLeuHisCysGluAspGlySerAspGluAlaAspCysValArgPhePhe	681	
Qy	568	GGA-----CCAAACTTTCATCTCTCCCTTCAGGTGTACTCATCTCTCAG	603	
Db	682	AsnGlyThrThrAsnAsnAsnGlyLeuValArgPheArgIleGln-----	696	
Qy	604	AGGAAGTCCTGGCACCCCTGTGTGCGAAGACGACTGGAAACGAGAACTAAGGGCGGCGGCC	663	
Db	697	---SerIleTrpHisThrAlaCysAlaGluAsnTrpThrThrGlnIleSerAsnAspVal	715	
Qy	664	TGCAGGACATGGGCTATAGAAATAATTTTACTCTAGCCCAAGGAATAGTGGATGACAGC	723	
Db	716	CysGlnLeuGlyLeuGlySerGly---AsnSerSerLysProIlePheSerThrAsp	734	
Qy	724	GGATCCACAGCTTTATGAATCAACAAAGTCGCGCAATTCGATATCTATAAAAA	783	
Db	735	GlyGly---ProPheValLysLeuAsnThrAlaPro-----AspGlyHisLeuIle	750	
Qy	784	CTGACACACAGTGCCTGTCTTCAAAACACAGTGTCTTCTTACCGTGT-----ATA	837	
Db	751	LeuThrProSerGlnGlnCysLeuGlnAspSerLeuIleArgLeuGlnCysAsnHisLys	770	
Qy	838	GCCTCGGGGTCAACTTGAATCTCAAGCCGC---CAGAGCAGGATTTGGGGGGGAGAGC	894	
Db	771	SerCysGlyLysLysLeuAlaGlnAspIleThrProLysIleValGlyGlySerAsn	790	
Qy	895	GGCTCCCGGGGCTGGCCCTGGCAGGTTCAGCTGCAGCTCCAGACGTCACGTTGCC	954	
Db	791	AlaLysGluGlyAlaIleProTrpValValGlyLeuTyTyGlyArgLeuLeuCys	810	
Qy	955	GGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCGGCCCTCGGTGGAAAAA	1014	
Db	811	GlyAlaSerLeuValSerSerAspTrpLeuValSerAlaAlaHisCysValTyArg	830	
Qy	1015	CTPAACAAATCCATGCGATTGACGGCATTCGGGGATTTTGAGACAACTCTTCATGTC	1074	
Db	831	AsnLeuGluProSerLysTrpThrAlaIleLeuGlyLeuHisMetLysSerAsnLeuThr	850	
Qy	1075	-----TATGGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCAAATATGATCTCC	1128	
Db	851	SerProGlnThrValProArgLeuIleAspGluIleValIleAsnProHisTyArgAsnArg	870	
Qy	1129	AAGACCAAGAACATGACATTCGCTGATGTAAGCTGCAGAGCCCTCTGACTTTCACGAC	1188	
Db	871	ArgArgLysAspAsnAspIleAlaMetMetHisLeuGluPheLysValAsnTyThrAsp	890	
Qy	1189	CTAGTGAACACAGTGTGTCTGCCCAACCCAGGCCATGATGTGTCAGCCAGCAACACTCTGC	1248	
Db	891	TyrIleGlnProIleCysLeuProGluGlnGlnValPheProProGlyArgAsnCys	910	
Qy	1249	TGGATTTCCGGGTGGGGGCCACCGAGGAAAGGAAAGACTCAGAAAGTCTGAAAGCT	1308	
Db	911	SerIleAlaGlyTrpGlyThrValValTyGlnGlyThrThrAlaAsnIleLeuGlnGlu	930	


```

Db      792 GlySerLeuIle-----LeuThrProSerLeuGlnCysSerGlnAspSerLeuIle 808
Qy      823 TCITTAGCTGT-----ATAGCTGGGGTCACTTGAACCTCAAGCGCCAGAGC--- 873
Db      809 LeuLeuGlnCysAsnHisLysSerCysGlyGluLysValThrGlnLysValSerPro 828
Qy      874 AGATTGTGGCGGAGAGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGAGCTCCAC 933
Db      829 LysIleValGlyGlySerAspAlaGlnAlaGlyAlaThrProTyrValValAlaLeuTyr 848
Qy      934 GTCAGAACGTCAC-----GTGTGGCGAGGCTCCATCATCATCCCCCGAGTGG 981
Db      849 HisArgAspArgSerThrAspArgLeuLeuCysGlyAlaSerLeuValSerAspIrp 869
Qy      982 ATCGTACAGCGCCCACTGGCTGGGAAACCTCTTAACATCATCGCATTCGACGGCA 1041
Db      869 LeuValSerAlaAlaHisCysValTyrArgArgAsnLeuAspProThrArgTyrThrAla 888
Qy      1042 TTTCGGGGATTGTGACACATCTTCACTGTCTCTATGAGCGCGATACCAA-----GTA 1095
Db      889 ValLeuGlyLeuHisMetGlnSerAsnLeuThrSerProGlnValValArgValVal 908
Qy      1096 GAAAGATGATTCTCTCAATTCATCTCCAGACTCCAGACCAAGAACAAATGACATTCGCTG 1155
Db      909 AspGlnIleValIleAsnProHisTyrAspArgArgArgLysValAsnAspIleAlaMet 928
Qy      1156 ATGAGCTGCGAAGCTCTGACTTTCACAGCACTAGTGAACACAGTGTGTCTGCCAAC 1215
Db      929 MetHisLeuGluPheLysValAsnTyrThrAspTyrIleGlnProIleCysLeuProGlu 948
Qy      1216 CAGGATGATGTGTCAGCAGACAGCTCTGTGTGATTTCGGGTGGGGGCCACCGAG 1275
Db      949 GluAsnGlnIlePheIleProGlyArgThrCysSerIleAlaGlyTyrGlyTyrAspLys 968
Qy      1276 ---GAGAAAGGAAGACTCAAGAGTGTCTGAAGCTCCAGAGGTCTTCTCATGAGACA 1332
Db      969 IleAsnAlaGlySerThrValAspValLeuLysGluAlaAspValProLeuIleSerAsn 988
Qy      1333 CAGAGATGACACAGACATGTCTATGACAACTGATCATCACAGCATGATCTGTGCC 1392
Db      989 GluLysCysGlnGlnLeuProGluTyrAsn---IleThrGluSerMetIleCysAla 1007
Qy      1393 GGCTTCTCGAGGGGAACGTGATCTTCCAGCGGTGCAGAGTGGAGCGCTCTGTGCTACT 1452
Db      1008 GlyTyrGluGluGlyIleAspSerCysGlnGlyAspSerGlyGlyProLeuMetCys 1027
Qy      1453 TCGAAGAACATATCTGGTGGCTGATAGGGATACAGCTGGGCTTGGCTGTGCCAAA 1512
Db      1028 GlnGluAsnAsnArgTyrPheLeuValGlyValThrSerPheGlyValGlnCysAlaLeu 1047
Qy      1513 GCTTACAGACCGAGGTGTACGGGAATGTGATGTATTTCAGCGACTGCATTTAT 1566
Db      1048 ProAsnHisProGlyValTyrValArgValSerGlnPheIleGluTyrPheHis 1065

RESULT 12
ENTK_PIG
ID ENTK_PIG STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Enteropetidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Ductal mucosa;
RX MEDLINE=94327548; PubMed=8051081;
RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S.,

```

```

RA Miki K., Kurokawa K., Tashiro K., Shiokawa K., Shinomiya K.,
RA Uneyama H., Inoue H., Takahashi T., Takahashi K.;
RT "Structural characterization of porcine enteropetidase.";
RL J Biol Chem. 269:19976-19982(1994).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterotrimer of a catalytic (light) chain, a multidomain
CC (heavy) chain, and a mini chain.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D30799; BAA06459.1; -.
CC HSSP: P00763; LDPO.
CC MEROPS: S01.156; -.
CC InterPro: IPR000859; CUB.
CC InterPro: IPR009003; Cys Ser trypsin.
CC InterPro: IPR002172; LDL receptor_A.
CC InterPro: IPR000998; MAM domain.
CC InterPro: IPR001254; Peptidase S1.
CC InterPro: IPR001344; Peptidase_S1A.
CC InterPro: IPR000082; SEA domain.
CC InterPro: IPR001190; Srcr_receptor.
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF00057; ldi_recept_a; 2.
CC Pfam: PF00629; MAM; 1.
CC Pfam: PF01390; SEA; 1.
CC Pfam: PF00330; SRCR; 1.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00261; CHYMOTRYPSIN.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PRINTS: PR00020; MAMDOMAIN.
CC SMART: SM00042; CUB; 2.
CC SMART: SM00192; LDLA; 2.
CC SMART: SM00137; MAM; 1.
CC SMART: SM00200; SEA; 1.
CC SMART: SM00202; SR; 1.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS01180; CUB; 2.
CC PROSITE: PS01209; LDLRA_1; 2.
CC PROSITE: PS00668; LDLRA_2; 2.
CC PROSITE: PS00740; MAM_1; 1.
CC PROSITE: PS00660; MAM_2; 1.
CC PROSITE: PS0024; SEA; 1.
CC PROSITE: PS00420; SRCR_1; FALSE_NEG.
CC PROSITE: PS0287; SRCR_2; 1.
CC PROSITE: PS0240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
KW Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease;
KW Zymogen; Transmembrane; Repeat; Lipoprotein.
FT CHAIN 52 117 NON-CATALYTIC M CHAIN (MINI CHAIN).
FT CHAIN 118 799 NON-CATALYTIC H CHAIN (HEAVY CHAIN).

```

FT	CHAIN	800	1034	CATALYTIC L CHAIN (LIGHT CHAIN).
FT	DOMAIN	1	18	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	19	47	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT				(POTENTIAL).
FT	DOMAIN	48	1034	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	52	169	SEA.
FT	DOMAIN	197	238	LDL-RECEPTOR CLASS A 1.
FT	DOMAIN	240	349	CUB 1.
FT	DOMAIN	357	519	NAM.
FT	DOMAIN	539	649	CUB 2.
FT	DOMAIN	656	694	LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	693	786	SER.
FT	DOMAIN	800	1034	SERINE PROTEASE.
FT	ACT SITE	840		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT SITE	891		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT SITE	986		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	N-myristoyl glycine (Potential).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	236	BY SIMILARITY.
FT	DISULFID	658	670	BY SIMILARITY.
FT	DISULFID	665	683	BY SIMILARITY.
FT	DISULFID	677	692	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	787	911	BY SIMILARITY.
FT	DISULFID	825	841	BY SIMILARITY.
FT	DISULFID	925	992	BY SIMILARITY.
FT	DISULFID	956	971	BY SIMILARITY.
FT	DISULFID	982	1010	BY SIMILARITY.
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	170	170	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	194	194	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	283	283	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	343	343	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	403	403	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	455	455	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	518	518	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	549	549	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	697	697	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	701	701	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	740	740	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	804	804	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	863	863	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1034 AA; 114776 MW; 038864CF64CC368 CRC64;		

Alignment Scores:

Pred. No.:	7, 67e-38	Length:	1034
Score:	634.00	Matches:	147
Percent Similarity:	49.58%	Conservative:	72
Best Local Similarity:	33.28%	Mismatches:	179
Query Match:	19.94%	Indels:	44
DB:	1	Gaps:	13

US-09-323-597c-1 (1-1738) x ENT_K_PIG (1-1034)

QY	304	GTCTGACGACGCCCAATCCCATCCGGGACAGTGTGCACCTCAAGACTAAGAACCA	363
DB	610	ValTyrThrGlyProGlyProValGluAspValPheSerThrThrAsnArgMetThrVal	629
QY	364	CTGTGCATCACCC-----TTGACCCCTGGGG-----ACCTTCCTCGTGGGA	402
DB	630	LeuPheIleThrAsnAspAlaLeuThrLysGlyGlyPheLysAlaAsnPheThrThrGly	649
QY	403	GCTGCGTGGCGCTGGGCTACTCTGTGAAGTTCATGGGACGACCAAGTGTCCCACTCTGGG	462
DB	650	TyrHisLeuGly-----11eProGluProCysLysGluAspAsn	662

QY	463	ATAGAGTGGCACTCCTCAGGTACCTGCATCAACCCCTCTAATCGTGTGTGATGGGTGTCA	522
DB	663	PheGlnCysGlu---AsnGlyGluCysValLeuLeuValAsnLeuCysAspGlyPheSer	681
QY	523	CACCTGCCCCGGGGAGGAGCAGAAATCGGTGTGTTCCTCTCTACGGA-----CCAAC	576
DB	682	HisCysLysAspGlySerAspGluAlaHisCysValArgPheLeuAsnGlyThrAlaAsn	701
QY	577	TTCATCTCTCAGGTGTACTCATCTCAGAGGAAGCTCTGGCACCCCTGTGTGCAAGACAC	636
DB	702	AsnSerGlyLeuValGlnPheArgGlnSerIleThrAlaCysAlaGluAsn	721
QY	637	TGGAACGAGAACTACGGGGCGGCGCTGCAGGACATGGGCTATAGAATAAT-----	690
DB	722	TrpThrThrGlnThrSerAspValCysGlnLeuLeuGlyLeuGlyThrGlyAsnSer	741
QY	691	-----TTTTACTCTAGCCCAAGGAATAGTGTGATGACAGCGGATCCACCGCTTTATG	741
DB	742	SerMetProPhePheSer-----SerGlyGlyGlyProPheVal	754
QY	742	AACTGACACACAGTCCCGCAATGTGATATCTATAAAACATGTACCAAGTATGCC	801
DB	755	LysLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrAlaSerGlnGln	771
QY	802	TGTTCTTCAAAAGCAGTGTTCCTTCTTACGCTGT-----ATAGCCTGCGGGTCAACTG	855
DB	772	CysPheGluAspSerLeuIleLeuGlnCysAsnHisLysSerCysGlyLysGln	791
QY	856	AACTCAAGCCCGCAGAC---AGGATTGTGGGGCGGAGCGCGCTCCCGGGGCGCTGG	912
DB	792	ValAlaGlnGluValSerProLysIleValGlyGlyAsnAspSerArgGluGlyAlaTrp	811
QY	913	CCCTGGCAGCTCAGCTGCAGTCCAGAACGTCACGTCGTGTGGAGGCTCCATCATCACC	972
DB	812	ProTrpValValAlaLeuTyrTyrAsnGlyGlnLeuLeuCysGlyAlaSerLeuValSer	831
QY	973	CCCGAGTGTGATCGTGACAGCGCCCACTGGGTGGAAAAACCTCTTAAACAATCCATGGCAT	1032
DB	832	ArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnLeuGluProSerLys	851
QY	1033	TGGACGGCATTTGGCGGATTTTGAGACAATCTTCATGTTCTTATGAGCGCGGATACAA	1092
DB	852	TrpLysAlaIleLeuGlyLeuHisMetThrSerAsnLeu-----ThrSerProGln	868
QY	1093	-----GTAGAAAAAGTGTATCTCATCCAAATATGATCCAAAGACCAACG	1137
DB	869	IleValThrArgLeuIleAspGluIleValIleAsnProHisTyrAsnArgArgLys	888
QY	1138	AACAATGACATTTGGCTGTGATGAAGCTGCAGAGCCTCTGACTTTTCAACGACCTAGTGA	1197
DB	889	AspSerAspIleAlaMetMetHisLeuGluPheLysValAsnTyrThrAspTyrIleGln	908
QY	1198	CCAGTGTGTCTGCCCAACCCAGCATGCTGCAGCCAGAACAGCTCTCTGATTTCC	1257
DB	909	ProIleCysLeuProGluGluAsnGlnValPheProProGlyArgIleCysSerIleAla	928
QY	1258	GGGTGGGGGCCCCAGGAGGAAGGAGACCTCAGAACTGTGTGAACGCTGCCAAGGTG	1317
DB	929	GlyTrpGlyLysValIleTyrGlnGlySerProAlaAspIleLeuGlnGluAlaAspVal	948
QY	1318	CTTCTCATTGACACACAGACATGCACAGACAGATATGCTATGACACACCTGATCACACA	1377
DB	949	ProLeuLeuSerAsnGluLysCysGlnGlnMetProGluTyrAsn-----IleThrGlu	967
QY	1378	GCCATGATCTGTCCGGCTTCCTCCAGGGGAAGCTCGATTCTTCCAGGTGTGACGTGA	1437
DB	968	AsnMetMetCysAlaGlyTyrGluGluGlyGlyIleAspSerCysGlnGlyAspSerGly	987
QY	1438	GGGCTTGTCTACTTCGAGAACATATCTGTGGCTGTAGTGGATACAGCTCGGGT	1497
DB	988	GlyProLeuMetCysLeuGluAsnAspArgTrpLeuLeuAlaGlyValThrSerPheGly	1007

Pred. No.:	1,79e-35	Length:	1042
Score:	601.00	Matches:	138
Percent Similarity:	51.15%	Conservative:	62
Best local Similarity:	35.29%	Mismatches:	169
Query Match:	18.91%	Indels:	22
DB:	1	Gaps:	14

US-09-323-597C-1 (1-1738) x CORI_HUMAN (1-1042)

QY	442	AGCAAGTCTCCAACTCTGGGATGAGAGCGCACTCTCCAGGTACTCATCTCAGAGGAAGTCC	501
DB	653	SerPheCysGlnAspAspGluLeuGluCys---AlaAsnHisAlaCysValSerArgAsp	671
QY	502	AACGTGCTGTGATGGCGTGTACACTGCCCCCGCGGAGAGACAGAGATCGGTGTGTCGC	561
DB	672	LeuTrpCysAspGlyGluAlaAspCysSerAspSerSerAspGluTrpAspCysValThr	691
QY	562	CTC-----TACGGACCAAACTTCATCTCAGGTGACTCATCTCAGAGGAAGTCC	612
DB	692	LeuSerIleAsnValAsnSerSerPheLeuMetValHisArgAlaAlaThrGlu---	710
QY	613	TGGCACTCTGTGTGCAAGACGACTGGAAACGAGAACTACGGCGCGGCGCTCGCAGGAC	672
DB	711	--HisHisValCysAlaAspGlyTrpGlnGluIleLeuSerGlnLeuAlaCysGln	729
QY	673	ATGGCTATAGATAATTTTACTCTAGCCAAAGATAGTGGATGACAGCGGATCCACC	732
DB	730	MetGlyLeuGlyGluProSerValThrLysLeu---IleGlnGluGlnGluLysGluPro	748
QY	733	AGCTTTATGAACCTGAACACA-----AGTCCGGCAATGTGATATCTATAAAAAA	783
DB	749	ArgTrpLeuLeuHisSerAsnTrpGluSerLeuAsnGlyThrLeuHisGluLeu	768
QY	784	CTGTACCACTGATGCTCTGTTTCAAAAGACGTGGTTCCTTAACGTGTATAGCC---	840
DB	769	LeuValAsnGlyGlnSerCysGluSerArgSerLysIleSerLeuLeuCysThrLysGln	788
QY	841	---TCGGGGTCAACTCACTCAAGCCGACGAGGATGTGGCGGCGAGGCGCG	897
DB	789	AspCysGlyArgProAlaAlaArgMetAsnLysArgIleLeuGlyGlyArgThrSer	808
QY	898	CTCCGGGGGCGTGGCCCTGGCAGGTCACTGACGTCAGCAG---AACGTCACGTGTC	954
DB	809	ArgProGlyArgTrpTrpTrpGlnCysSerLeuGlnSerGluProSerGlyHisIleCys	828
QY	955	GGAGCTCCATCATACCCCGAGTGGATCTGACAGCGCCGACCGCCACTGGTGGAACCT	1014
DB	829	GlyCysValLeuIleAlaLysLysTrpValLeuThrValAlaHisCysPheGlu---	847
QY	1015	CTTAACAATCCATGCGATTCGACGCGCATTCGCGGCAAT-----TTGACAGCAATCTTC	1068
DB	848	ArgGluAsnAlaAlaValTrpLysValValLeuGlyIleAsnAsnLeuAspHisProSer	867
QY	1069	ATGTTCTATGGACGCGGATACCAAGTAGAAAAGTATTTCTTCFCCAATATGACTCC	1128
DB	868	ValPheMetGlnThrArgPhe---ValLysThrIleLeuHisProArgTrpSerArg	886
QY	1129	AAGACCAAGAACATGACATTCGCTGATGAGCTGACAGAGCCTCTCAGCTTCAACGAC	1188
DB	887	AlaValValAspTrpAspIleSerIleValGluLeuSerGluAspIleSerGluThrGly	906
QY	1189	CTAGTAAACCACTGTGTCTCCCAACCCAGCATGATGCTGACGACGAGACAGCTGTC	1248
DB	907	TyrValArgProValCysLeuProAsnProGluGlnTrpLeuGluProAspThrTyrCys	926
QY	1249	TGGATTCCGGGTGGGGGCCACCGAGGAGAAAGACCTCAGAGAGTGTGCAACGCT	1308
DB	927	TyrIleThrGlyTrpGly-----HisMetGlyAsnLysMetProPheLysLeuGlnGlu	944
QY	1309	GCCAAAGTCTCTCATTTGACACACAGAGATGCAACAGCAGATATGTCTATGCAACCTG	1368
DB	945	GlyGluValArgIleIleSerLeuGluHisCysGlnSer---TyrPheAspMetLysThr	963

RESULT 14

CORI_MOUSE

ID CORI_MOUSE STANDARD; PRT; 1113 AA.

AC Q92319;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-converting enzyme) (Corin) (low density lipoprotein receptor related protein 4).

DE protein 4).

GN CRN OR LRP4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

XX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96429596; PubMed=9756524;

RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;

RT "A novel low-density lipoprotein receptor-related protein with type II membrane protein-like structure is abundant in heart.";

RL J. Biochem. 124:784-789 (1998).

CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Type II membrane protein.

CC -1- TISSUE SPECIFICITY: Highly expressed in heart.

CC -1- SIMILARITY: Belongs to peptidase family S1.

CC -1- SIMILARITY: Contains 2 frizzled (FZ) domains.

CC -1- SIMILARITY: Contains 7 LDL-receptor class A domains.

CC -1- SIMILARITY: Contains 1 SRCR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB013874; BAA34371.1; -

PIR; J0315; J0315.

HSP; P00763; IDPO.

MEROPS; S01.019; -

 MGD; MGI.1349451; LRP4. | InterPro; IPR009003; Cys_Ser_trypsin. | InterPro; IPR000024; Fz_domain. | InterPro; IPR002172; LDL_receptor_A. | InterPro; IPR001254; Peptidase_S1. | InterPro; IPR001314; Peptidase_S1A. | InterPro; IPR001190; Srcr_receptor. | Pfam; PF001392; Fz; 2. | Pfam; PF00057; ldl_recept_a; 6. | Pfam; PF00089; trypsin; 1. | PRINTS; PR00722; CHYMOTRYPSIN. | PRINTS; PR00261; LDLRECEPTOR. | SMART; SM00063; FRI; 2. | SMART; SM00192; LDLa; 3. |

760	LeuSerLysAsnGlyAsnSerSerSerLeuLeuThrValHisLysSerAlaLysGlu---	778
613	TGGCACCCCTGTGTGCAAGACAGACTGGGAACAGAACTACGGCGCGGGCGCTCGAGGAC	672
779	---HisHisValCysAlaaspGlyTrrArgGluThrLeuSerGlnLeuAlaCysLysGln	797
673	ATGGCGTATAGAATAATTTTACTCTACCAAGGAATAGTGGATGACACGGGATCCACC	732
798	MetGlyLeuGlyGluPro-----SerValThrLysLeuIleProGlyGlnGluGlyGln	815
733	AGCTTTTAAACTCAACCAAGTCGCGCAATCTCGAT-----ATCTATAAAAA	793
816	GlnTrpLeuArgLeuTyrrProAsnTrpGluAsnLeuAsnGlySerThrLeuGlnGluLeu	835
784	CTGTACCAACAGNATGCGCTGTTCTTCAAAACAGCATGGTTCTTTACCTGTATAGCC---	840
836	LeuValTyrrArgHisSerCysProSerArgSerGluIleSerLeuLeuCysSerLysGln	855
841	---TGCGGGGTCAACTCAACTCAAGCGCCAGACGAGGATTGTGGCGCGCAGAGCGCG	897
856	AspCysGlyArgArgProAlaAlaArgMetAsnLysArgIleLeuGlyArgThrSer	875
898	CTCCCGGGGCGCTGCGCTGCGAGGTGAGCTGCAGTCCAG-----AACGTCCAGTGTGC	954
876	ArgProGlyArgTrpTrpTrpGlnCysSerLeuGlnSerGluProSerGlyHisIleCys	895
955	GGAGGCTCCATCATCACCCCGAGGTGATCGTACACGCGCCCACTCGGTGGAA-----	1008
896	GlyCysValLeuIleAlaLysLysTrpValLeuThrValAlaHisCysPheGluGlyArg	915
1009	-----AAACCTCTTAAATCCATGCCAT	1032
916	GluAspAlaAspValTrpLysValValPheGlyIleAsnAsnLeuAspHisPro-----	933
1033	TGACGGCATTTGCGGGATTTTGAGACAATCTTTCATGTTCTATGGAGCGGATACCA	1092
934	-----SerGlyPheMetGlnThrArgPhe-----	941
1093	GTAGAAAGATGATTTCTCATCCAAATATCACTCCCAAGACCAGAAACAATGACATGGC	1152
942	ValLysThrIleLeuLeuHisProArgTyrrSerArgAlaValAlaAspTyrrAspIleSer	961
1153	CTGATGAGCTGCAGAACCTCTCACTTTCAACGACCTAGTGAACACGAGTGTCTGCCC	1212
962	ValValGluLeuSerAspAspIleAsnGluThrSerTyrValArgProValCysLeuPro	981
1213	AACCCAGGATGATGCTCTCAGCAGAACAGCTCTGCTGCGATTTCGGGTGGGGCGCAC	1272
982	SerProGluGluTyrrLeuGluProAspThrTyrrCysTyrrIleThrGlyTrpGly----	999
1273	GAGGAGAAAGGAAGACCTCGAAGTGTGAACCTGCCAAGGTGCTTCTCATTTAGACA	1332
1000	HisMetGlyAsnLysMetProPheLysLeuGlnGluGlyValArgIleIleProLeu	1019
1333	CAGAGATCAACAGCAGATATGTCTATGACAACTGATCACACGCGCATGATCTGTGCC	1392
1020	GluGlnCysGlnSer---TyrrPheAspMetLysThrIleThrAsnArgMetIleCysAla	1038
1393	GCGTCTTCGACGGGAACGTCAATCTTTCGACGGGTGACAGTGGAGGCGCTCTGGTCACT	1452
1039	GlyTyrrGluSerGlyThrValAspSerCysMetGlyAspSerGlyGlyProLeuValCys	1058
1453	TCGAAG---AACATATCTGCTGCTGATAGGGGATACAGCTGGGTTCTGCGCTGT---	1506
1059	GluArgProGlyGlyGlnTrpThrLeuPheGlyLeuThrSerTrpGlySerValCysPhe	1078
1507	GCCAAAGCTTACAGACCAAGGAGTGTACGGGAATGTGATGTATTTCACGGACTGGATTAT	1566
1079	SerLysValLeuGlyProGlyValTyrrSerAsnValSerTyrrPheValGlyTrpIleGlu	1098
1567	CGACAAATGAGGCGAGACGGCTAATCCACATGCTCTGCTGCTGAGTCGCTTTTACAG	1626
1099	ArgGlnIleTyrrIleGln-----ThrrPheLeuGlnLys	1109

Alignment Scores:	
Pred. NO.:	6.88e-34
Score:	579.00
Percent Similarity:	48.58%
Best Local Similarity:	32.55%
Query Match:	18.31%
DB:	1
	Gaps: 15
	Indels: 56
	Mismatches: 162
	Conservative: 68
	Matches: 138
	Length: 1113

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:54:20 ; Search time 84 Seconds
(without alignments)
11692.085 Million cell updates/sec

Title: US-09-323-597C-1
Perfect score: 3179
Sequence: 1 ggcggaggcggaggcggagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues
Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+72p.model -DEV=xlp
-Q=/cgs2_1/USPTO.spool_P/US09323597/runat 06072004 111727 29633/app query.fasta_1.1927
-DB=A Geneseq 29Jan04 -OPMT=fastan -SUPFIX=tag -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptn -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597@cgs2_1_81@runat 06072004 111727 29633 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	85.5	492	3 AAY44406	Aay44406 Human 20P
2	2717	85.5	492	3 AAY77726	Aay77726 Human tum
3	2717	85.5	492	5 AAE18096	Aae18096 Human 20P
4	2713	85.3	492	5 AAE18098	Aae18098 Human 20P
5	2713	85.3	492	5 AAE18100	Aae18100 Human 20P
6	2713	85.3	492	5 AAE18099	Aae18099 Human 20P
7	2713	85.3	492	5 AAE18101	Aae18101 Human 20P
8	2711	85.3	492	7 ADB75591	Adb75591 Prostate
9	2709	85.2	492	3 AAY92050	Aay92050 HrpCag/7
10	2709	85.2	492	4 AAM01315	Aam01315 P1000C am

11	2709	85.2	492	4 AAY69960	Aay69960 Human pro
12	2709	85.2	492	4 ABU71851	Abu71851 Prostate
13	2709	85.2	492	5 ABB95420	Abb95420 Human P10
14	2709	85.2	492	6 ABR54532	AbR54532 Prostate
15	2709	85.2	492	7 ADB14382	AdB14382 Human pro
16	2696	84.8	492	3 AAY57280	Aay57280 Ovr115 ho
17	2696	84.8	492	3 AAB36901	Aab36901 Human TMP
18	2696	84.8	492	4 AAM01294	Aam01294 Human tra
19	2696	84.8	492	4 AAU69939	Aau69939 Human tra
20	2696	84.8	492	4 AAE06943	Aae06943 Human ser
21	2696	84.8	492	4 ABU71830	Abu71830 Prostate
22	2696	84.8	492	5 ABB95399	Abb95399 Human tra
23	2696	84.8	492	5 ABG61885	Abg61885 Prostate
24	2696	84.8	492	5 AAE18097	Aae18097 Human TMP
25	2696	84.8	492	6 ABR48241	AbR48241 Human bla
26	2696	84.8	492	6 ABR54511	AbR54511 Human tra
27	2696	84.8	492	7 ADB14345	AdB14345 Human tra
28	2171	68.3	393	4 AAM01317	Aam01317 P1000C pa
29	2171	68.3	393	4 AAU69962	Aau69962 Human pro
30	2171	68.3	393	4 ABU71853	Abu71853 Prostate
31	2171	68.3	393	5 ABB95422	Abb95422 Human P10
32	2171	68.3	393	6 ABR54534	AbR54534 Prostate
33	2171	68.3	393	7 ADB14384	AdB14384 Human pro
34	2126	66.9	384	6 ADA10905	Ada10905 Human CDN
35	1540	48.4	283	3 AAY81492	Aay81492 Human pro
36	1540	48.4	283	5 ABB78328	Abb78328 Amino aci
37	1540	48.4	283	5 AAE14615	Aae14615 Human tra
38	1165	36.6	209	4 AAM01295	Aam01295 Human tra
39	1165	36.6	209	4 AAU69940	Aau69940 Human tra
40	1165	36.6	209	4 ABU71831	Abu71831 Prostate
41	1165	36.6	209	5 ABB95400	Abb95400 Human tra
42	1165	36.6	209	6 ABR54512	AbR54512 Partial H
43	1165	36.6	209	7 ADB14347	AdB14347 Human tra
44	1019.5	32.1	418	6 ABB56845	Abb56845 Consensus
45	888	27.9	454	3 AAB32246	Aab32246 Tumour as

ALIGNMENTS

RESULT 1
AAY44406
ID AAY44406 standard; protein; 492 AA.
XX
AC AAY44406;

DT 22-MAR-2000 (first entry)

DE Human 20P1F12-GTC2 protein.

KW 20P1F12; TWPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;
transmembrane protein; colon; prostate; prostate tumour.

XX Homo sapiens.

XX WO9962942-A2.

XX 09-DEC-1999.

XX 01-JUN-1999; 99WO-US012253.

XX 01-JUN-1998; 98US-0087598P.

XX 14-APR-1999; 99US-0129521P.

XX (UROC-) UROGENESYS INC.

XX (AFAP/) AFAP D E.

XX (HUBE/) HUBERT R S.

XX (LEON/) LEONG K.

XX (RAIT/) RAITANO A B.

XX (SAFF/) SAFFRAN D C.

XX Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;

DR WPI; 2000-116363/10.
 XX N-9SDB; AAZ229636.

PI Novel cell surface antigen useful to treat colon and prostate cancer.

PS Claim 1; Fig 1; 58pp; English.

XX The present sequence is the 20p1f12 protein (also known as the TMPRSS2
 CC protein) which is a prostate-specific, androgen-regulated, cell surface
 CC serine protease. It is a glycosylated type II transmembrane protein with
 CC an extracellular C-terminal serine protease domain, a scavenger receptor
 CC cysteine-rich domain, an LDL receptor class A domain and a predicted
 CC transmembrane domain. Host cells can be transformed to produce this
 CC protein, using vector containing 20p1f12/TMPRSS2 gene (also designated
 CC 20p1f12-GTC1, as deposited with ATCC accession number 207097). Anti-
 CC 20p1f12/TMPRSS2 antibodies may be used as therapeutic agent for prostate
 CC and colon cancers, to image prostate cancer cells and prostate tumours,
 CC to identify ligands and cellular constituents that bind to a
 CC 20p1f12/TMPRSS2 gene product and for use as cancer vaccines

XX Sequence 492 AA;

Alignment Scores:

2red. No.: 7,27e-222 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 3 Gaps: 0

US-09-323-597C-1 (1-1738) x AAY44406 (1-492)

QY 112 ATGGCTTTGAACACTAGGCTACCCAGCTATTGACCTTACTATGAAACCAATGATAC 171
 DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAACCCCTATCCCGCAGCCACTGCTGCTCCACCTGCTACGAGTGCAT 231
 DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACTACCCGTCGCCCGCCAGTACGCCCGAGGCTCCTCAGCAGGCT 291
 DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCGCTGCTGACCGCAGCCCAATCCCATCCGCGACAGTGTGCACTCAAG 351
 DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCTCGGCGACCTTCTCGTGGAGCTCGGCTG 411
 DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 QY 412 GCGGTGGCCTACTCTGGAAGTTTCATGGGAGCAAGTCTCCTCACTCTGGGATAGTGC 471
 DB 101 AlaAlaGlyLeuLeuTrpPheMetGlySerLysCysSerAsnSerGlyLeuGluCys 120
 QY 472 GACTCTCAGGTACCTGCATCAACCCCTCACTGCTGTGATGGGTGTCACTGCCCC 531
 DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGGAGGAGGAGAGATGGTGTTCGCTCTACGACCAAACTTCATCTTCAGGTG 591
 DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGAGAGTCTGGCACCTGTGTGCAAGACGACTGGAAACGAGAACTAC 651
 DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGGGCGGCTGCGAGGACATGGGCTATAGAAATAATTTTACTTAGCAAGAAATA 711
 DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGACGGGATCCACGAGCTTTATGAACCTGAACACAGTCCCGGCAATGTGAT 771

DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAATAAACTGTACACAGTGTCTCTTCAAAACAGCTGTTCTTTACGC 831
 DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTTGGCGGCTCAACTTGAACCTCAAGCCGCGCAGACGAGGATTTGGCGCGCGAG 891
 DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGGCTCCCGGGCGCTGGCCCTGGAGCTCAGCTGCACCTCCAGAACGTCACGCTG 951
 DB 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TGGCGAGGCTCCATCATCACCCCGAGTGCATCGTCACAGCCGCCACTCGCTGGGAAAAA 1011
 DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTAACAATCCATGGCATTCGAGCGGCAATTTGGGGAGTTTGTAGACAACTTTCATG 1071
 DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGGAGCCGCGATACCAAGTAGAAAAAGTCAATTTCTCATCAAAATATGACTCCAG 1131
 DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGAACAAATCAGATTGGCTGTGATGAAGCTGCAGAGCCTCTGACTTTCAACGCTTA 1191
 DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGluLysProLeuThrPheAsnAspLeu 360
 QY 1192 GTCAACACCATGCTGTCTGCCCAACCCAGGCATCATCTGCAGCCAGAACAGCTCTGCTGG 1251
 DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 QY 1252 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGGAAGACCTCAGAGTGTCTGAACGCTGCC 1311
 DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 QY 1312 AAGTGCTTCTCATTTGACACACAGATGCAACAGAGATATCTCTATGACCACTGATC 1371
 DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACCAAGCATGATGTGTGCGCGCTTCTGCAGGGGAAACGTCGATTTCTGCCAGGGTGAC 1431
 DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGCGCTCTGTGCTACTTCGAAGAACAAATATCTGTGGCTGTAGGGGATACAGC 1491
 DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACGAGGAGTGTACGGGAATGTGATGGTATTC 1551
 DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 QY 1552 ACAGACTCGGATTTTATCGACAAATGAGCGGACAGCGC 1587
 DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2

AAAY77726
 ID AAAY77726 standard; protein; 492 AA.

XX AC AAY77726;

XX 12-MAY-2000 (first entry)

XX Human tumour suppressor TMPRSS2 polypeptide.

XX Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
 KW gene therapy; protein therapy.

XX Homo sapiens.

XX W020000605-A1.
 XX PD 06-JAN-2000.
 XX PF 29-JUN-1999; 99WO-US014622.
 XX PR 29-JUN-1998; 98US-0091044P.
 XX PA (MYRI-) MYRIAD GENETICS INC.
 XX PI Wong AKC, Tavtigian SV, Teng DHF;
 XX DR WPI; 2000-170914/15.
 XX DR N-PSDB; AA287786.
 XX PT Novel tumor suppressor TMRPS2 used for the diagnosis and prognosis of
 XX PT human cancer.
 XX PS Claim 55; Page 77-79; 89pp; English.
 XX CC The invention provides a new tumor suppressor gene, designated TMRPS2.
 CC The TMRPS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with anti
 CC -cancer or therapeutic properties. The polypeptides are also useful for
 CC rational drug design. The TMRPS2 polynucleotides and polypeptides may be
 CC used for gene therapy and protein therapy. The present sequence
 CC represents the TMRPS2 polypeptide
 XX
 XX SQ Sequence 492 AA;

Alignment Scores:
 Pred. No.: 7,27e-222 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 3 Gaps: 0

US-09-323-597C-1 (1-1738) x AAV77726 (1-492)

QY 112 ATGGCTTGAATCAGGCTGACACACAGCTATTGGACCTTACTATGAAACACGATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAAACCCCTATCCCGACAGCCCACTGFGTCCCACTGCTCTACGAGGTGCAT 231
 Db 21 GluProGluAsnProTyrProAlaGluProThrValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACACCGCTCCCGCTGCCAGTACGCCCCGAGGGTCTCTGACGAGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCCTGCTGTCAGCGAGCCAAATCCCAATCCCGGACAGTGTGCACCTCAAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCTTCACCTCGGACCTTCTCTGCGGAGCTGCGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 CCGCTGCTCTCTGGAAGTTCATGGCGAGGAGTCTCCAACTGCGGATAGAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCAGGTACTCTCATCAACCCCTCTAACTGGTGTGATGCGGTGTACACTGCCCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 QY 532 GCGCGGAGACGAGAAATCGGTGTGCTTCCGCTCTACGGACCAACTTCATCTTCAGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160

RESULT 3
 AA218096
 ID AA218096 standard; protein; 492 AA.
 XX

QY 592 TACTCATCTCAGAGGAAGTCTCTGGCAACCTGTGTGCTCCAGACGACTGGAGCACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
 QY 652 GGGCGCGGCTCTGAGGAGACATGGCTATAAGATAATATTTTACTCTAGCCAAAGATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAACTGCGCGGATCTCGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACAGTGTCTCTTCAAAAGCAGTGTCTTTCTTTTACGC 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTTGGGGGTCAACTTGAACCTCAAGCCCGCAGAGCAGAGTGTGGCGCGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGGCTCCCGGGGCGCTGGCCCTGGCAGGTGACCTGCAGTCCAGAACCTCCACGTG 951
 Db 261 SerAlaLeuProGlyAlaIleTrpTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TCGGAGGCTCCATCATCACCCCCGAGTGTGACAGCGCGCCCTGCTGCTGGAATAA 1011
 Db 281 CysGlyGlySerIleIleThrProGluThrIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTACATCCATGGCATGGACCGCATTTGGGGGATTTTGAGACAACTCTTTCATG 1071
 Db 301 ProLeuAsnAsnProTrpHisIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGGAGCCGGATACCAAGTAGAAAAGTATTCTTCATCCAAATATGACTCCAAG 1131
 Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGAACATGCAATTGCGCTGTAGAGCTGCGAGAGCCCTGCACTTTCACAGACCTA 1191
 Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysThrSerGluValLeuAsnAlaAla 360
 QY 1192 GTGAACACCATGTGTCTGCCAACCCAGGCATGTGTGTCAGCCAGAACACAGCTCTCTCG 1251
 Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 QY 1252 ATTTCCGGGTGGGGGCGCCACCGAGAGAAAGGAGAGACCTCAGAGTGTCTGACCGTGC 1311
 Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 QY 1312 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
 Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACCAAGCCATGATCTGTCCCGCTTCTCGAGGGAAGCTGCTGATCTTTCAGGAGTAC 1431
 Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGCTCTGCTGCTCACTTTCGAGACATATCTGCTGCTGCTGATGCGGATACAAG 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIlePheLeuIleGlyAspThrSer 460
 QY 1492 TGGGGTCTGCTGCTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTGTTTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 QY 1552 AGGACTGATTTATCGACAAATGAGGCGCAGCGGC 1587
 Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

AAE18096;
 07-MAY-2002 (first entry)
 Human 20P1F12-GTCl protein.
 Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCl; cell growth; neoplasm;
 cancer; vaccine; human.
 Homo sapiens.
 WO200204953-A2.
 17-JAN-2002.
 12-JUL-2001; 2001WO-US022168.
 12-JUL-2000; 2000US-00615295.
 (AGEN-) AGENSYS INC.
 Safarzan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
 Challita-Bid PW;
 WPI; 2002-154967/20.
 N-PSDB; AAD28778.
 Examining a biological sample for evidence of dysregulated cellular
 growth, comprises comparing the status of prostate-specific, androgen-
 regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
 normal sample.
 Claim 40; Fig 1; 161pp; English.
 The present invention relates to methods and compositions for the
 diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
 kidney cancer derived from or based on a normally prostate-specific,
 androgen regulated, cell membrane associated secreted serine protease
 termed 20P1F12/TMPRSS2. The invention further relates to a method of
 examining a biological sample for evidence of dysregulated cellular
 growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
 designated 20P1F12-GTCl) in the sample to the status of 20P1F12/TMPRSS2
 in a corresponding normal sample. The invention also relates to 20P1F12/
 TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
 invention are used for examining a sample such as blood, serum, stool,
 urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
 The dysregulated cell growth is indicative of bladder cancer, lung
 cancer, kidney cancer or ovarian cancer. It is useful for identifying
 evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
 portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
 expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
 lung cancer, ovarian cancer or metastatic cancer. The present sequence is
 human 20P1F12-GTCl protein

Alignment Scores:
 Pred. No.: 7,278-222 Length: 492
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.47% Indels: 0
 DB: 5 Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18096 (1-492)

QY 112 ATGGCTTGAATCAGGCTCACCACGCTATTGACCTTACTATGAACCACTGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAAAACCCCTATCCCGCACAGCCCACTGTGTCCCGACCTGTCTACAGATGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40

QY 232 CGGGCTCAGTACTACCCGTCCTCCCGTCCCGAGTACGCCCGAGGGTCTCTGACGAGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCGCTGCTCTGACGAGCCCAATCCCATCCGAGACAGTGTGCACCTCAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAAGCACTGTGCATCACCTTGACCTCGGGGACCTTCTCTCGTGGAGCTGGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGCTGGCTACTCTGGAAGTTCATGGGAGCAAGTGTCTCCAACTCTGGGATAGAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCTGTGTACACTGCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGGGAGACGAGAAATCGTGTGTCTTACGGACCAAACTTCATCTCTTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGAGGAAGTCTTGGCACCTGTGTCACAGACGACAGTGGACGAGAACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGCGGCGGCTCTGACGAGGACATGGGCTATAAGAATAATTTTCTCTTACCCAGGAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATCAGCGGATCCACAGCTTATGAACCTGAACACAGTCCCGCAATGTCGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACAGTGTGCTGTCTTCTTCAAAAGCAGTGGTTCCTTACGC 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTTGGGGGTCACTTGAACCTGAGCCGCGCAGAGCAGGATTTGGGCGGCGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 951
 Db 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TGGGAGGCTCCATCATCACCCCGAGTGTGATCGTGACAGCCCGCCCTCGCTGGGAAAAA 1011
 Db 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTAACAATCCATGGCATTGCGGCGATTTCGGGGATTGTGAGACAATCTTTTCATG 1071
 Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGGAGCCGGATACCAAGTAGAAAAAGTAGTTTCTCATCCAAATTATGACTCCAAG 1131
 Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGACATACATTTGGCTGATGACAGCTGCAGAGGCTCTGACTTTTCAACGACCTA 1191
 Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 QY 1192 GTGAACCAAGTGTGTCTGCCCCAACCCAGGATGATGTGACGACGAGCAACAGCTCTGCTGG 1251
 Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 QY 1252 ATTTCCGGTGGGGGGCCACCAGGAGAAAGGAAGCACTCAGAGTGTGAGAGCTGCC 1311
 Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400

QY 1312 AAGGTGCTTCTATTGACACAGAGATGCAACAGCAGATATCTCTATGACAACTGATC 1371
 Db 401 LysValLeuLeuLeuGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACCGCCATGATCTGCGCGCTCTCCAGCGGAACCTCGATTCTTCCAGGGTGAC 1431
 Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGGCTCTGGTCACTTCGAAGAACAAATATCTGGTGTATAGGGGATACAAGC 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGTCTCTGCTGTGCGCAAGCTTACAGACAGGAGTGTACGGAAATGTGATGTTATTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 QY 1552 ACGGACTGGATTTATCGACAAATGAGGCGCAGACGCGC 1587
 Db 481 ThrAspTrpIleTyrArgGlnMetMetArgAlaAspGly 492

RESULT 4

AAE18098
 ID AAE18098 standard; protein; 492 AA.

AC AAE18098;

XX 07-MAY-2002 (first entry)

XX Human 20P1F12/TMPRSS2 mutant protein (S441A).

XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
 cancer; vaccine; human.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 441 /note= "Wild type Ser substituted with Ala"

XX W0200204953-A2.

XX 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US022168.

XX 12-JUL-2000; 2000US-00615285.

XX (AGEN-) AGENSYS INC.

XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
 Challita-Bid PM;

DR WPI; 2002-154967/20.

XX Examining a biological sample for evidence of dysregulated cellular
 growth, comprises comparing the status of prostate-specific, androgen-
 regulated, secreted serine protease, 20P1F12/TMPRSS2, in a correspond-
 ing normal sample.

XX Example 10; Page; 161pp; English.

XX The present invention relates to methods and compositions for the
 diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
 kidney cancer derived from or based on a normally prostate-specific,
 androgen regulated, cell membrane associated secreted serine protease
 termed 20P1F12/TMPRSS2. The invention further relates to a method of
 examining a biological sample for evidence of dysregulated cellular
 growth comprising comparing the status of 20P1F12/TMPRSS2 gene (also
 designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2
 in a corresponding normal sample. The invention also relates to 20P1F12/
 TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
 invention are used for examining a sample such as blood, serum, stool,

CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
 CC The dysregulated cell growth is indicative of bladder cancer, lung
 cancer, kidney cancer or ovarian cancer. It is useful for identifying
 evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
 portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
 expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
 lung cancer, ovarian cancer or metastatic cancer. The present sequence is
 human 20P1F12/TMPRSS2 mutant protein (S441A). Note: This sequence is not
 shown in the specification but is derived from wild type 20P1F12/TMPRSS2
 protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18098)

XX Sequence 492 AA;

Alignment Scores:
 Pred. No.: 1,31e-221 Length: 492
 Score: 2714.00 Matches: 491
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.80% Mismatches: 0
 Query Match: 85.37% Indels: 0
 DB: 5 Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18098 (1-492)

QY 112 ATGGCTTTGAACCTCAGGCTCACCACGACTATTGGACCTTACTATGAAACCATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACCGGAAACCCCTATCCCGCAGCCACCTGTCGCCCTCTCTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACTACCCGTCGCCCGTCCCGTCCAGTACGCCCGAGGTCCTGACGCGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAACCCGTCGTCTGCGACGAGCCCAATCCCATCCGCGGACGTCGTCCTCAAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCCTGTGCATCACCTTGACCTCTGGGACCTTCTCTGTGGGAGCTGCGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGCTGGCTACTCTGGAAAGTTTCATGGCAGCAAGTCTCCACTCTGGATAGAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCCTCAGTACCTGCATCAACCCCTCTAACTGTGTGATGGCGTGTCACTGCGCTG 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGCGGAGCAGAGAAATCGTGTGTTCGCTCTACGACCCAACTTCATCTCTGAGGTC 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGGAGTCTCGGACCCCTGTGTCGAAGACGACGACGACGACGACGACGAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GCGCGGCGGCTCTGACGAGGACATGGCTATAGAATAATTTTCTCTCTACCCAGGAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGGATGACGCGGATCCACGAGCTTTATGAACCTGAACACAGTCCCGCATGTGCTAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACACAGTGTGCTGTTCTTCAAAAGACAGTGGTTCCTTACGCG 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGATATAGCTTGGCGGTCAACTTGAATCACTCAAGCCCGCAGAGGAGTTGTGGCGGCGAG 891

Db 241 CysIleAlaCysGlyValAlaAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Qy 892 AGCGGCTCCCGGGGCTGGCCCTGGCAGGTGACGCTGACGCTCCAGACGTCCAGTG 951
Db 261 SerAlaLeuProGlyAlaTrpProGlnValSerLeuHisValGlnAsnValHisVal 280
Qy 952 TGGGAGGCTCCATCATCACCCTCCGAGTGGATCTGTACAGCGGCCACTGGTGAAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Qy 1012 CCTCTTAAATCCATGATGACGCGCATTTGGGGGATTTTGAGACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Qy 1072 TTTATGAGCGGATACCAAGTAGAAAAGTATTTCTATCCAAATATGATCCAG 1131
Db 321 PheTyGlyAlaGlyTyGlnValGluLysValIleSerHisProAsnTyAspSerLys 340
Qy 1132 ACCAAGAACATGACATTCGCTGATGAGCTGCAGAGCTCTGACTTTCAAGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Qy 1192 GTGAACACGAGTGTCTCCCAACCCAGGCGATGATGCTGCAGGCCAGACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Qy 1252 ATTTCCGGGTGGGGGCGCACCGAGGAGAGGAGACCTCAGAGTCTGAGCTGAGCTGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Qy 1312 AAGGTGCTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1371
Db 401 LysValLeuLeuLeuLeuThrGlnArgCysAsnSerArgTyValTyAspAsnLeuIle 420
Qy 1372 ACACAGCATGATCTGTGCGGCTCTTCGAGGGAAGCTGATCTTGCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Qy 1432 AGTGAGGCGCTCTGCTCACTTCGAGAACATATCTGTGTGCTGATAGGGGATACAAAGC 1491
Db 441 AlaGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Qy 1492 TGGGTTCTGCTGCGCAAGCTTACAGACAGAGTGTACGGGAATGTATGATATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyArgLysValMetValPhe 480
Qy 1552 ACGGACTGATTTATCGACAAATGAGGCGAGACGGC 1587
Db 481 ThrAspTrpIleTyArgGlnMetArgAlaAspGly 492

RESULT 5
AAE18100
ID AAE18100 standard; protein; 492 AA.
XX
AC AAE18100;
XX
DT 07-MAY-2002 (first entry)
XX
DE Human 20P1F12/TMPRSS2 mutant protein (R252Q).
XX
KW Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCL; cell growth; neoplasm;
KW cancer; vaccine; human.
XX
OS Homo sapiens.
OS Synthetic.
PH
FT Key Location/Qualifiers
FT Misc-difference 252 /note= "wild type Arg substituted with Gln"
XX
FN W0200204953-A2.
XX
PD 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US022168.
XX
XX 12-JUL-2000; 2000US-00615285.
XX
PA (AGEN-) AGENSYS INC.
XX
XX Saferan D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
PI Challita-Bid PM;
XX
XX WPI; 2002-154967/20.
XX
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX
XX Example 10; Page: 161pp; English.
XX
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTCL) in the sample to the status of 20P1F12/TMPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12/TMPRSS2 mutant protein (R252Q). Note: This sequence is not
XX shown in the specification but is derived from wild type 20P1F12/TMPRSS2
XX protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
XX Sequence 492 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1.59e-221 Length: 492
XX Score: 2713.00 Matches: 491
XX Percent Similarity: 100.00% Conservative: 1
XX Best Local Similarity: 99.80% Mismatches: 0
XX Query Match: 85.34% Indels: 0
XX DB: 5 Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18100 (1-492)

Qy 112 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyTyTyGluAsnHisGlyTy 20
Qy 172 CAACCGAAACCCCTATCCCGACACGCCCTGCTGCTGCCACTGTCTACGAGTGCTAT 231
Db 21 GlnProGluAsnProTyProAlaGlnProThrValProThrValTyGluValHis 40
Qy 232 CCGCTCAGTACTACCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
Db 41 ProAlaGlnTyTyTyProSerProValProGlnTyAlaProArgValLeuThrGlnAla 60
Qy 292 TCCAAACCCCGTGTCTGACGAGCCCAATCCCATCCCGGACAGTGTGTCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Qy 352 ACTAAGAAACGACGTGTCATCCTTCACCTCGGGGACCTTCTCTCGGGGAGCTGCCTG 411
Db 81 ThrLysAlaLeuLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100

QY 412 GCCGCTGGCTACTCTGGAAGTTTCATGGCCAGCAAGTCTCCAACTCTGGATAGATGTC 471
Db 101 AlaAlaGly/LeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCTCAACCCCTAACTGCTGTGATGGGTGTCTACACGTGCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGGAGGACGAGAATCGGTGTGTCCTCTACGGACCAAACTTCATCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAGTCTGGACCCCTGTGTCGCAAGCAGTCCGAGCAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyr 180
QY 652 GCGCGGGCGCTGAGGACATGGCTATAAGAAATAATTTTACTAGCCCAAGCAATA 711
Db 181 GlyArgAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAAAGTCCCGGCAATGCT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACACAGTATGCTGTCTTCAAAAGCAGTGGTTCCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTGGCGGCTCAACTGAACCTCAAGCCGCGCAGCAGGATTTGGCGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerGlnGlnSerArgIleValGlyGlu 260
QY 892 AGCGGCTCCCGGGCGCTGGCCGCGCAGGTCCAGCTCCAGTCCAGAACCTCCACGTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCCGCGCCACTCGCTGGA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CTTCTTAACAATCCATGATGGACGGCATTTGGCGGATTTTGACAACTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTTATGAGCGCGATACCAAGTAGAAAAGTATTTCTATCCAAATATGACTCCAA 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGACAAATGACATTTGCGTGATGAGCTGCAGAACCTCTGACTTTCACAGCTTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACAGTGTCTGCCACCCAGCAGCATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGCCCCGAGGAGAAAGGAAGACCTCAGAACTGTCTGACCGTGC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTCTCTTCTATGACACACAGATCCACAGCAGATATGCTATGACACACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAAGCATGATCTGTCCGGCTTCTCCAGGGAAGCTCGATTTCTGCCAGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlnValAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGAGGGGCTCTGGTCACTTCGAAACAATATCTGTGGTGTAGGGGATACAGC 1491
Db 441 SerGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGGCTGTGCAAGCTTACAGACAGGAGTGTACCGGAATGATGGTATTC 1551

Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTATCGACAAATCAGGCGCAGACCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
RESULT 6
AAE18099
ID AAE18099 standard; protein; 492 AA.
XX
AC AAE18099;
XX 07-MAY-2002 (first entry)
DE Human 20P1F12/TMPRSS2 mutant protein (R240Q).
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
KW cancer; vaccine; human.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 240 /note= "Wild type Arg substituted with Gln"
XX
PN WO200204953-A2.
XX
PD 17-JAN-2002.
XX
PF 12-JUL-2001; 2001WO-US022168.
XX
PR 12-JUL-2000; 2000US-00615285.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Saferran D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
PI Chailita-Bid BW;
XX
DR WPI; 2002-154967/20.

Examining a biological sample for evidence of dysregulated cellular growth, comprises comparing the status of prostate-specific, androgen-regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding normal sample.

Example 10; Page; 161pp; English.

The present invention relates to methods and compositions for the diagnosis and therapy of prostate, colon, bladder, lung, ovarian and kidney cancer derived from or based on a normally prostate-specific, androgen regulated, cell membrane associated secreted serine protease termed 20P1F12/TMPRSS2. The invention further relates to a method of examining a biological sample for evidence of dysregulated cellular growth comprising comparing the status of 20P1F12/TMPRSS2 gene (also designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2 in a corresponding normal sample. The invention also relates to 20P1F12/TMPRSS2 polynucleotides and their corresponding proteins. Methods of the invention are used for examining a sample such as blood, serum, stool, urine, semen, or biopsy tissue for evidence of dysregulated cell growth. The dysregulated cell growth is indicative of bladder cancer, lung cancer, kidney cancer or ovarian cancer. It is useful for identifying evidence of a neoplasm in a sample. Vaccines comprising an immunogenic portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer, lung cancer, ovarian cancer or metastatic cancer. The present sequence is human 20P1F12/TMPRSS2 mutant protein (R240Q). Note: This sequence is not shown in the specification but is derived from wild type 20P1F12/TMPRSS2 protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE180996) Sequence 492 AA;

Alignment Scores:

Pred. No.:	1.59e-221	Length:	492
Score:	2713.00	Matches:	491
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.80%	Mismatches:	0
Query Match:	85.34%	Indels:	0
DB:	5	Gaps:	0

US-09-323-597C-1 (1-1738) x AAE18099 (1-492)

112	QY	ATGGCTTTGAACCTCAGGCTCAGCAGCAGCTATTGGACCTTACTATGAAAACCATGGATAC	171
1	Db	1 MetaAlaLeuAenSerGlySerProProAlaIleGlyProTyTyroGluAenHisGlyTy	20
172	QY	CAACCGGAAAAACCCCTATCCCGCAGACGCCACACTGTGGTCCCACTCTACACAGGTGCAT	231
21	Db	21 GlnProGluAenProTyProAlaGlnProThrValValProThrValTyroGluValHis	40
232	QY	CGGGCTCAGTACTACCGTCCCGCGTCCCGCAGTACGCCCGCAGGGTCTCGACGAGCGCT	291
41	Db	41 ProAlaGlnTyTyProSerProValProGlnTyAlaProArgValLeuThrGlnAla	60
292	QY	TCCAAACCCCGTCTCTGCACGACGCCAAATCCCCATCCGGGACAGTGCACCTCAAAG	351
61	Db	61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
352	QY	ACTAAGAAGCACTGTGCATCACCCTGCCTGGGACCTTCCTCTGGGGAGCTCGCGTG	411
81	Db	81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
412	QY	GCGCTGCGCTACTCTGTGAAGTTCATGGGACGAGTGTCCAACTCTCGGATAGAGTGC	471
101	Db	101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
472	QY	GACTCCTCAGGTACTGCATCAACCCCTCTAACTGGTGTGATGGCGTGCACACTGCCCC	531
121	Db	121 AspSerSerGlyThrCysIleAenProSerAsnTrpCysAspGlyValSerHisCysPro	140
532	QY	GCGGGGAGGACGAACTCGGTGTTCGGCTCTAGGACCAACTTCATCTCTTCAGCTG	591
141	Db	141 GlyGlyGluAspGluAenA-GcysValArgLeuTyroGlyProAenPheIleLeuGlnVal	160
592	QY	TACTCATCTCAGAGGAAGTCTCGCACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC	651
161	Db	161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAenTy	180
652	QY	GGCGGGCGGCTCGACGGACATGGCTTAAGAATAATTTTACTCTAGCCCAAGGAATA	711
181	Db	181 GlyArgAlaAlaCysArgAspMetGlyTyroLysAsnAenPheTyroSerSerGlnGlyIle	200
712	QY	GTGGATGACAGCGGATCCACGAGCTTTATGAAACTGAACACAGTCCCGGCAATGTCCAT	771
201	Db	201 ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp	220
772	QY	ATCTATAAAACCTGTACCACTGATGCCTGTCTTTCAAAAGCAGTGTTCCTTTAGCG	831
221	Db	221 IeTyroLysLysLeuTyroHisSerAspAlaCysSerSerLysAlaValValSerLeuGln	240
832	QY	TGTATAGCTCCGGGGTCAACTTCAACTCAAGCGCCGACAGCAGGATTTGTGGCGCGCAG	891
241	Db	241 CysIleAlaCysGlyValAenLeuAenSerSerArgGlnSerArgIleValGlyGlu	260
892	QY	AGCGGCTCCCGGGGCGTGGCCCTCGCAGGTGAGCTGCAGTCCGACAACTCCACGCTG	951
261	Db	261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAenValHisVal	280
952	QY	TGCGAGAGCTCCATCATCCCCCGAGTGGATCGTACAGCGCCCACTGCGTGGGAAAA	1011
281	Db	281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys	300
1012	QY	CCTCTTAACATCCATGGCATTTGGACGGCATTTGCGGGGATTTTGAGACAAATCTTTCATG	1071
301	Db	301 ProLeuAenAenProTrpHisIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320

QY	1072	TTCTATGGAGCCCGGATACCAAGTAGAAAGATGATTTCTCATCCCAATATGACTCCAAG	1133
Db	321	PhetyrGlyAlaGlyrGlnValGluLeuValIleSerHisProenlyrAspSerLys	340
QY	1132	ACCAAGAAACAATGACATTGGCTGTGATGAAGCTGCAGAAGCCTCTGACTTTTCAACGACCTA	1191
Db	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
QY	1192	GTGAAACCAAGTGCTGTGCCCAACCCAGGACATGATGCTCCAGCCAGACAGCTCTGCTGG	1251
Db	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
QY	1252	ATTTCGGGGGTGGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAGTGTGAACGCTGCC	1311
Db	381	IleSerGlyrTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
QY	1312	AAAGTGGCTCTCAATCAGACACAGAGATCCAAACAGCAGATATCTCTATGACACCTGATC	1371
Db	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTrpValTrpAspAsnLeuIle	420
QY	1372	ACACCAGCCANGATGTGTCCGGCTTCCTGCGAGGAAACGTGATTTCTGCCAGGGGTGAC	1431
Db	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
QY	1432	AGTGGAGGGCTCTGTGCTCACTTCGAGACACAAATATCTGTGGCTGATAGGGATACAAAGC	1491
Db	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer	460
QY	1492	TGGGGTCTTGCTGTGCTCAAAAGCTTACACACCAGGAGTGTACGGGAATGTGATGGTATTTC	1551
Db	461	TrpGlySerGlyCysAlaLysAlaTrpArgProGlyValTrpGlyAsnValMetValPhe	480
QY	1552	ACGGATCGATTATTCGACAAATAGAGGGCAGACGGC	1587
Db	481	ThrAspTrpIleTrpArgGlnMetArgAlaAspGly	492
RESULT 7			
AAE18101			
ID	AAE18101 standard; protein; 492 AA.		
XX	AAE18101;		
XX	07-MAY-2002 (first entry)		
XX	Human 20P1F12/TMPRSS2 mutant protein (R255Q).		
DE	Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCl; cell growth; neoplasm;		
KW	cancer; vaccine; human.		
KW	Homo sapiens.		
OS	Synthetic.		
OS	Key		
XX	Location/Qualifiers		
PH	Key		
FT	Misc-difference 255		
FT	/note= "Wild type Arg substituted with Gln"		
XX	WO200204953-A2.		
PN	17-JAN-2002.		
XX	12-JUL-2001; 2001WO-US022168.		
PF	12-JUL-2000; 2000US-00615285.		
XX	(AGEN-) AGENSYS INC.		
XX	Saferan D, Raitano AB, Hubert RS, Jakobovits A, Paris M;		
PI	Challita-Sid FM;		
XX	WPI; 2002-154967/20.		
XX	Examining a biological sample for evidence of disregulated cellular		

Examining a biological sample for evidence of dysregulated cellular

growth, comprises comparing the status of prostate-specific, androgen-regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding normal sample.

Example 10; Page; 161pp; English.

The present invention relates to methods and compositions for the diagnosis and therapy of prostate, colon, bladder, lung, ovarian and kidney cancer derived from or based on a normally prostate-specific, androgen regulated, cell membrane associated secreted serine protease termed 20P1F12/TMPRSS2. The invention further relates to a method of examining a biological sample for evidence of dysregulated cellular growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also designated 20P1F12-GT1) in the sample to the status of 20P1F12/TMPRSS2 in a corresponding normal sample. The invention also relates to 20P1F12/TMPRSS2 polynucleotides and their corresponding proteins. Methods of the invention are used for examining a sample such as blood, serum, stool, urine, semen, or biopsy tissue for evidence of dysregulated cell growth. The dysregulated cell growth is indicative of bladder cancer, lung cancer, kidney cancer or ovarian cancer. It is useful for identifying evidence of a neoplasm in a sample. Vaccines comprising an immunogenic portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer, lung cancer, ovarian cancer or metastatic cancer. The present sequence is human 20P1F12/TMPRSS2 mutant protein (R255Q). Note: This sequence is not shown in the specification but is derived from wild type 20P1F12/TMPRSS2 protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)

Sequence 492 AA;

Alignment Scores:

Pred. No.: 1.59e-221 Length: 492
Score: 2713.00 Matches: 491
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.80% Mismatches: 0
Query Match: 85.34% Indels: 0
DB: Gaps: 0

US-09-323-597C-1 (1-1738) x AAE18101 (1-492)

112 ATGGCTTTGAAGTCAAGGTCACACAGCTATTGGACTTACTATGAAACCACTGATAC 171
1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
172 CAACCGGAAACCCCTATCCGACACAGCCCACTGTGTCTCCCACTGTCTACGAGTGCAT 231
21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
232 CCGGCTCAGTACTACCCGTCCTCCCGTCCCGCAGTACGCCCGGGTCTCGACCGAGCT 291
41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
292 TCCAAACCCCTGCTGTCACGCGCCCAATCCCGATCCGGACAGTGTGCACCTCAAG 351
61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
352 ACTAAGAAACACTGTGCATCCTGACCTCGGACCTTCTCTGTTGGAGCTGGCTG 411
81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
412 GCGCTGCTACTCTGGAAGTTTCATGGGAGCAAGTCTCCAACTCTGGATAGAGTGC 471
101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
472 GACTCTCCTCAGTACCTGCATCAACCCCTCTAATGTTGTGATCGGCTGTCACACTGCC 531
121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
532 GGGGGGAGGACGAGAAATCGGTGTGTTTCGCTCTACGACCAAACTTCATCTTCAGGTG 591
141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
592 TACTCATCTCAGAGGAGTCTCGCACCTGTGTGCCAAGACGACTGGAGACGAGACTAC 651

161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
652 GGGGGGCGGCTGTCAGGAGCATGGGCTATATAAGATAATTTTACTCTACCCAGGAATA 711
181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
712 GTGGATGACAGCGGATCCACAGCTTTATGAAACTGAACACACAGTCCCGGCAATGTCTGAT 771
201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
772 ATCTATAAAAACTGTACACAGTGTGCTCTTCTTCAAAAGCAGTGTGTTCTTTACGC 831
221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
832 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCCGACAGCAGGATTCGTGGCGCGCAG 891
241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerGlnIleValGlyGlyGlu 260
892 AGCGGCTCCCGGGGCGCTGGCCCTGGCAGGTGAGCTGCACGTCCAGAACGTCACGTG 951
261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
952 TCGGAGGCTCCATCATCATCCCGGAGTGGATGGTGCACAGCCGCCCTGCGTGGGAAA 1011
281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
1012 CTTCTTAACAAATCCATGGATTCGAGCGCATTTGGCGGATTTTGGAGACAATCTTTCATG 1071
301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
1072 TTCTATGGAGCGGATACCAAGTAGAAAAGTGAATTTCTATCCCAATATGACTCAAG 1131
321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
1132 ACCAAGAACATGACATTCGCTGATGAAGCTGCAGAGCCCTCTGACTTTCAACGACCTA 1191
341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
1192 GTGAACCAAGTGTGTGCTGCCAACCCAGCATGATGCTGCAGCCAGAACAGCTCTGCTGG 1251
361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
1252 ATTTCCGGTGGGGGCCACCCAGGAGAAAGGAGACCTCAGAGTGTCTGAACGCTGCC 1311
381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
1312 AAGGTGCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
1372 ACACGAGCATCATCTGTGCGGCTTCTGAGGGGAGAGCTCGATTTCTGCCAGGTGAC 1431
421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
1432 AGTGGAGGCTCTGTCTACTTCGAGAACATATCTGTGCTGATAGGGATACAAGC 1491
441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
1492 TGGGTTCTGCTGTGCCAACTTACAGACAGAGGTGACGGGAATGTGATGTTATTC 1551
461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
1552 ACGGACTGATTTATCGACAAATGAGGCGCAGCGC 1587
481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 8

ADB75591

ID ADB75591 standard; protein; 492 AA.

XX

AC ADB75591;

XX

DT 04-DEC-2003 (first entry)
 XX Prostate cancer marker protein.
 XX Prostate; cancer; cytostatic; gene therapy; marker.
 XX Homo sapiens.
 XX WC2003009814-A2.
 XX 06-FEB-2003.
 XX 25-JUL-2002; 2002WO-US023913.
 XX 25-JUL-2001; 2001US-0307982P.
 PR 22-AUG-2001; 2001US-0314356P.
 PR 25-SEP-2001; 2001US-0325020P.
 PR 12-DEC-2001; 2001US-0341746P.
 PR 05-MAR-2002; 2002US-0362158P.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;
 PI Hoersht S, Kamatkar S, Wonsey AM, Glatt K, Zhao X, Anderson D;
 XX WPI; 2003-248033/24.
 XX New nucleic acid molecule, useful for diagnosing or treating prostate cancer.
 PT Disclosure; SEQ ID NO 415; 99pp; English.
 XX The invention relates to newly discovered cancer markers associated with the cancerous state of prostate cells. Also disclosed is a method of assessing whether a patient is afflicted with prostate cancer. The method of the invention involves assessing whether a patient is afflicted with prostate cancer by comparing the level of expression of a marker in a patient sample and the normal level of expression of the marker in a control non-prostate cancer sample, where a significant increase in the level of expression of the marker in the patient sample and the normal level indicates that the patient is afflicted with prostate cancer.
 CC Nucleic acids of the invention are useful for diagnosing or treating prostate cancer, and may be useful in gene therapy. Sequences given in ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 492 AA;

Alignment Scores:
 Pred. No.: 2,35e-221 Length: 492
 Score: 2711.00 Matches: 490
 Percent Similarity: 100.00% Conservative: 2
 Best Local Similarity: 99.53% Mismatches: 0
 Query Match: 85.28% Indels: 0
 DB: 7 Gaps: 0

US-09-323-597C-1 (1-1738) x ADB75591 (1-492)

QY 112 ATGGCTTTGAACTCAGGTCCACACAGCTATTGGACCTTACTATGAACCATGGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 QY 172 CAACGGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 QY 232 CCGGCTCAGTACTACCGTCCCGCTGCGCCAGTACGGCCGAGGGTCTCTGACGAGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProGlnValLeuThrGlnAla 60
 QY 292 TCCAAACCCCGTCTGTGCACGCGACGCCAAATCCCCATCCGGGACAGTGTGCACCTCAAG 351

Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCCTTGACCTCGGGAGCTTCTCTGCTGGGAGCTCGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 QY 412 GCCCTCGCCTACTCTGAAAGTTCAVGGGAGCAAGTGTCTCAACTCTGGGATAGATGTC 471
 Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCAGGTACTGCATCAACCCCTCTAACTGTGTGTGATGGCTGTCTCATCTGCCCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGCGAGGACGAGAATCGGTGTGCTCTACGACCAAACTTCTCATCTTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
 QY 592 TACTCATCTCAGAGAAAGTCTGTGCACCCCTGTGTGCGCAAGCAGTCTGGAACAGACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 QY 652 GGGGGGGGGCTGACGGACATGGCTATAAGAAATATTTTACTCTAGCAAGGAATA 711
 Db 181 GlyArgAlaAlaCysLysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 QY 712 GTGATGACAGCGGATCCACAGCTTTATGAACATGAACACAACTGCGGCAATGTGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAACCTGTACCAAGTACGCTGTCTTCAAAAGCAGTGTCTTCTTCACG 832
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
 QY 832 TGTATAGCCTCGCGGGTCAACTTGAACCTCAAGCCCGCAGAGCAGATTTGGGGCGGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 AGCGGCTCCGGGGGCTGGCCCTGCGAGTCAAGCTGCGACGTCAGACAGTCCAGACGTC 951
 Db 261 SerAlaLeuProGlyAlaIleProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TCGCGAGGCTCCATCATCACCCTCGAGTGGATCGTGACAGCCGCCCACTCGTGGAAAAA 1011
 Db 281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CTCTTAACTTCCATGGCATTTGACCGGCTTTCGGGGATTTTGACACAACTTTCATG 1071
 Db 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTGTCTCTCATCCAAATTTACTCCAAAG 1131
 Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 QY 1132 ACCAAGAACAAATGACATTTGCGTGTGATGAAGCTGACAGAGCTCTGTGCTTTCACAGCCTA 1191
 Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 QY 1192 GTGAACCCAGTGTCTCTGCCCAACCCAGGCTATGCTGTCAGCAGACAGCTCTGCTGG 1251
 Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 QY 1252 ATTTCCGGTGGGGGCCACCGGAGGAAAGGAAAGACCTCAGAGTGTCTGAGCTGCC 1311
 Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 QY 1312 AAGTGTCTCTCATTTGACACAGAGTGCACAGCAGATATGCTATGATGACACCTGATC 1371
 Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 QY 1372 ACACCCAGCATGATCTGTGCGGCTTCTGCGAGGGAACCTGATTTCTTCCAGGGGTGAC 1431

Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 QY 1432 AGTGGAGGCGCTCTGGTCACTTCGAGAACAAATATCTGGTGGCTGATAGCGGATACAAGC 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpLeuIleGlyAspThrSer 460
 QY 1492 TGGGGTTCTGGCTGTCGCAAGCTTACAGACCAAGAGTGTACGGGAATGTGATGGTATTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyArgLysAsnValMetValPhe 480
 QY 1552 ACGACTGGATTATTCGACAAATGAGCGGACGCGC 1587
 Db 481 ThrAspTrpIleTyArgGlnMetArgAlaAspGly 492

RESULT 9

AA92050
 ID AA92050 standard; protein; 492 AA.

XX AC AA92050;

XX DF 01-AUG-2000 (first entry)

XX DE HrPCa6/7 polypeptide from androgen-inducible gene clone.
 XX KW Androgen inducible; testosterone; prostate cancer; cytostatic; TMPRSS2;
 XX KW diagnosis.
 XX OS Homo sapiens.

XX PN W0200018961-A2.

XX PD 06-APR-2000.

XX PF 30-SEP-1999; 99WO-US022535.

XX PR 30-SEP-1998; 98US-00163759.

XX PR 30-SEP-1998; 98US-00164159.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Macbeth KJ, Shyjan AW;

XX DR WPI; 2000-293182/25.

XX DR N-PSDB; AAA08803.

XX PT Novel methods for identifying compounds for treating prostate cancer
 PT comprising measuring the level of expression or activity of 1 or more of
 PT 11 genes or their products.
 XX CS Claim 2; Fig 3; 108pp; English.

XX CC This protein is encoded by a gene which is androgen (e.g. testosterone)
 CC inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells)
 CC and constitutively expressed in androgen-independent prostate cancer
 CC cells (e.g. LNCap cells). Agents which decrease the expression or
 CC activity of these clones may slow or arrest the growth of prostate cancer
 CC cells or may kill them. HrPCa6/7 can be obtained from the sequence of the
 CC known gene for TMPRSS2. A compound useful for treating prostate cancer
 CC can be identified in a novel method comprising measuring the expression
 CC level, or activity, of HrPCa2, 3, 6/7, 8, 9, 10, 13, 14, 15, 19, or
 CC peripheral-type benzodiazepine receptor (PBR) in a cell, in the presence
 CC and absence of a test compound. The sequences may also be used in
 CC diagnosis of prostate cancer and to determine efficacy of treatment for
 CC prostate cancer

XX SQ Sequence 492 AA;

Alignment Scores:

Pred. No.: 3,48e-221 Length: 492
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservatives: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 85.22% Indels: 0

DB: 3 Gaps: 0
 US-09-323-597C-1 (1-1738) x AA92050 (1-492)
 QY 112 ATGGCTTTGCACTCAGGTTCACCAACAGCTATTGGACCTTACTATGAAAAACCATGATAC 171
 Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyTrpGluAsnHsGlyTy 20
 QY 172 CAACCGGAAAAACCCCTATCCCGCAGCCCACTGTGGTCCCACTGTCTTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyTrpAlaGlnProThrValValProThrValTyTrpGluValHis 40
 QY 232 CCGGCTCAGTACTACCCGTCGCCCTGCCCCAGTACGCCCGAGGGTCTTCAGCAGGCT 291
 Db 41 ProAlaGlnTyTrpProSerProValProGlnTyTrpAlaProArgValLeuThrGlnAla 60
 QY 292 TCCAAACCCCGTCTGTGCAGCAGCCCAATCCCATCCGACAGTGTGCACCTCAAAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCCCTGGGAGCTTCTCTGTGGAGCTGGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 QY 412 GCGGCTGGCTTACTTGGAAAGTTCTGCGCAGCAAGTCTCCCACTCTGGGATAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 QY 472 GACTCTCTCAGGTACCTGCATCAACCCCTCTAACTGCTGTGTGATGGGTGTGCACCTGCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 QY 532 GCGGCGGAGGACAGAAATCGGTGTGTGCTTACGACCAACCACTTCTCTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyGlySerAsnPheIleLeuGlnVal 160
 QY 592 TACTCATCTCAGAGGAAGTCTGGCACCCTGTGTGTCAGACCACTGGAACCAAGCACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTy 180
 QY 652 GCGGCGGCGCTGCAGGACATGGGCTATAAGAATAATTTTCTTAGCCCAAGAAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyTrpLysAsnAsnPheTyTrpSerGlnGlyIle 200
 QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAAAGTCCGGCAATTCGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 QY 772 ATCTATAAAAACTGTACCACAGTATGCTGTTCTTCAAAAGCAGTGGTTTCTTTACGC 831
 Db 221 IleTyTrpLysLysLeuTyTrpHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 QY 832 TGTATAGCTGGGGGTCACTTGAACCTCAAGCCCGCAGAGCAGGATTTCTGGCGCGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 QY 892 ACGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGCAGCTGCAGCTCCAGAACCTCCACGTG 951
 Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 QY 952 TCGGAGGCTCCATCATATACCCCGAGTGGATGTGACAGCCCGCCACCTCGCTGGAAAAA 1011
 Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 QY 1012 CCTCTTAACAATCCATGGCATTTGGACCGCATTTGGGGGATTTTGAGACAATCTTTTCATG 1071
 Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 QY 1072 TTCTATGAGCGCGATACCAAGTAGAAAAAGTATTTCTCATCCAAATATGACTCCCAAG 1131
 Db 321 PheTyTrpGlyAlaGlyTyTrpGlnValGluLysValIleSerHisProAsnTyTrpAspSerLys 340
 QY 1132 ACCAAGAACAATGACATTGCGCTGATGAAGCTGCAGAGCCCTCTGACTTTTCAACGACCTA 1191

Db 341 ThrLysAsnAsnAspIleAlaLeuMethLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Qy 1192 GTGAACCCAGTGTCTGCCCAACCCAGGCGATGATGCTGGAGCCAGACAGCTCTGCTGG 1251
 Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 Qy 1252 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGAGCTCAGAAAGTGTCTGAAGCTGCTGCC 1311
 Db 381 IleSerGlyTrpGlyAlaThrGluGlnLysGlySerGlyValLeuAsnAlaAla 400
 Qy 1312 AAGTGTCTCTCATTGACACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
 Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Qy 1372 ACACAGCAGTATCTGTCGGCTCTCTGAGGAGACGTCGATCTTTCACGGGTGAC 1431
 Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Qy 1432 AGTGAGGGCTCTGTGTCACCTTCGAAGAACATATCTGTGGCTGATAGGGGATACAGC 1491
 Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 Qy 1492 TGGGTCTGCTGTCGCAAGCTTACAGACAGAGTGTACGGGAATGCTGATGCTATTC 1551
 Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Qy 1552 ACGGACTCGATTATCGACAAATGAGGGCAGACGGC 1587
 Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 10

AAM01315
 ID AAM01315 standard; protein; 492 AA.

AC AAM01315;

DT 04-OCT-2001 (first entry)

DE P1000C amino acid sequence.

KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;

KX Cystostatic; gene therapy; metastasis.

OS Homo sapiens.

PN WO200151633-A2.

PP 19-JUL-2001.

PR 16-JAN-2001; 2001WO-US001574.

PR 14-JAN-2000; 2000US-00483672.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SB, Jiang Y, Reed SG;

PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM;

PI Wang A, Meagher MJ;

XX WPI; 2001-425873/45.

XX New polynucleotide encoding a prostate-specific protein, for diagnosing,

PT monitoring and treating prostate cancer in a patient and for use in

PT vaccines.

XX Claim 2; Page 539-540; 543pp; English.

PS The present invention describes polynucleotide sequences (I) which encode

CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,

CC and can be used in vaccine production and gene therapy. (I), (II),

CC antibodies to (II), fusion proteins comprising (II), and isolated T cells

CC prepared using (I) or (II) are used treat cancer in a patient. (I) and

CC the antibodies are also used in the detection of cancer in a patient. The
 CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
 CC and (II) can be used in vaccines. The antibodies or (I) can be used for
 CC monitoring the progression of cancer in a patient. (I) and (II) can also
 CC be used to improve diagnostic and therapeutic methods for prostate
 CC cancer. They can indicate the level of metastasis as well as the prostate
 CC volume. AAH93357 to AAH93944 and AAM01115 to AAM01318 represent
 CC polynucleotide and amino acid sequences used in the exemplification of
 CC the present invention
 XX

SQ Sequence 492 AA;

Alignment Scores:

Pred. No.: 3,486-221 Length: 492
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 85.22% Indels: 0
 DB: Gaps: 0

US-09-323-597C-1 (1-1738) x AAM01315 (1-492)

Qy 112 ATGGCTTTGAACTCAGGTCACCCAGCCTATTGGACCTTACTATGAACCATGGATC 171
 Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Qy 172 CAACCGGAAACCCCTATCCCGCACGCCACTGTGTCTCCCACTGTCTACGAGGTGCAT 231
 Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 Qy 232 CCGGCTCAGTACTACCGTCCCGGTCGCCAGTACGCCCGGAGGTCCTGAGCGAGGCT 291
 Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Qy 292 TCCAAACCCCTCGTCTGCACGACGCCCAATCCCATCCGCGACAGTGTGCACCTCAAAG 351
 Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Qy 352 ACTAAGAAGCACTGTGCATCAGCTTACCTCGGGGACCTTCTCTGGGAGCTGGCTG 411
 Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Qy 412 GCGCTGGCCCTACTCTGGGAAGTTTCATGGGCGAGCAAGTGTCTCCAACTCTGGAGTAGAGTGC 471
 Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Qy 472 GACTCTCAGGTACCTGCATCAACCCCTCTAAGTGTGTGATCGGTCGCTGCACCTGCC 531
 Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Qy 532 GCGGGGAGGACGAGAAATCGGTGTTCGCTCTACGGACCAAACTTTCATCTTCAGGTG 591
 Db 141 GlyGlyGluAspGluAsnArgCysValAlaGLeuTyrGlySerAsnPheIleLeuGlnVal 160
 Qy 592 TACTCATCTCAGAGGAGTCTCGGCACCTGTGTGCCAAGACGACTGGACGAGAACTAC 651
 Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 Qy 652 GGGCGGGCGGCTGCGAGGACATGGCTATAGAATAATTTTACTCTAGCAAGGAATA 711
 Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 Qy 712 GTGATGACAGCGGATCCACCACTTTATGAACACTGAACAAAGTCGCGCAATGTGAT 771
 Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Qy 772 ATCTATATAAACTGTACCACTGATGCTGTCTTCAAACGAGTGGTTCCTTTACGC 831
 Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Qy 832 TGTATAGCTTCGGGGTCAACTTGAACACTCAAGCCCGCAGACGAGGATTTGGCGCGGAG 891
 Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260


```
QY 892 AGCGCGCTCCCGGGGCGCTGGCCCTGCGAGTGCAGCTGCAGCTCCAGAACTGCCAGTG 951
PR |||||
DB 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGAGTGCATCGTGCAGCGCCGCCACTGCGTGGAAAG 1011
DB |||||
DB 281 CysGlySerIleIleTrpGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTCGACGGCATTTGCGGGATTTTGAGACAATCTTTCATG 1071
DB |||||
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGACCGCATACCAAGTAGAAAAGTGAATTTCTCATCCAAATTAATGACTCAAG 1131
DB |||||
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTCGCTGATGAAGCTCAGAGAGCTCTGACTTTTCAACGACCTA 1191
DB |||||
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACCATGCTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTCTGG 1251
DB |||||
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAGTGTGAAACGCTGCC 1311
DB |||||
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGCTCTCTATGACACAGAGATGCAACAGCAGATATCTCTATGACAACTGTATC 1371
DB |||||
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGACCATGATCTGTGCGGCTTCCTGCAGGGAACGTGCAATCTGCCAGGCTGAC 1431
DB |||||
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGTGTCACCTTCGAGACAAATATCTGTGGCTGATAGGGGATACAAGC 1491
DB |||||
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGGCTGTGCGCAAGCTTACAGACCGAGTGTACGGGAATGTGATGTTATTC 1551
DB |||||
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGATTTATCGACAAATCAGGCGCAGACGGC 1587
DB |||||
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 11
AAU69960
XX AC AAU69960 standard; protein; 492 AA.
XX AC AAU69960;
XX AC AAU69960;
DT 30-JAN-2002 (first entry)
XX DE Human prostate cDNA encoded protein #86.
XX KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX OS Homo sapiens.
XX PN WO200173032-A2.
XX PD 04-OCT-2001.
XX PP 27-MAR-2001; 2001WO-US009919.
XX PR 27-MAR-2000; 2000US-00536857.
XX PR 09-MAY-2000; 2000US-00568100.
XX PR 12-MAY-2000; 2000US-00570737.
XX PR 13-JUN-2000; 2000US-00593793.
```

```
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter WM, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX WPI; 2001-639232/73.
DR N-PSDB; AAS64178, AAS64179, AAS64180.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
XX
PS Claim 2; Page 573-574; 579pp; English.
```

The invention relates to isolated prostate-specific polynucleotides, polypeptides, fusion proteins of the polypeptides, antibodies raised against the polypeptides (or antigenic epitopes derived from them) and antigen-presenting cells expressing the polypeptides. The antibodies are useful for detecting the presence of cancer, especially prostate cancer. The polypeptides, polynucleotides and the antigen-presenting cells are useful for stimulating and/or expanding T cells specific for a tumour protein, and for inhibiting the development of cancer especially prostate cancer. Compositions comprising the polynucleotide and/or polypeptide are useful for stimulating an immune response, and for treating cancer. The oligonucleotide is useful for detecting cancer. The present sequence is a prostate specific polypeptide of the invention

Sequence 492 AA;

Alignment Scores:

Pred. No.:	3,48e-221	Length:	492
Score:	2709.00	Matches:	491
Percent Similarity:	99.80%	Conservative:	0
Best Local Similarity:	99.80%	Mismatches:	1
Query Match:	99.82%	Indels:	0
DB:	4	Gaps:	0

US-09-323-597C-1 (1-1738) x AAU69960 (1-492)

QY	112	ATGGCTTTGAACCTCAGGCTCACCACGCTATTGGACCTTACTATGAAACCATGGATAC	171
DB	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrIleGluAsnHisGlyTyr	20
QY	172	CAACCGGAAACCCCTATCCCGCAGCCCACTGTGGTCCCCCACTGTCTPACGAGTGCAAT	231
DB	21	GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
QY	232	CGGGCTCAGTACTACCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT	291
DB	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
QY	292	TCCAAACCCCGTGTCTGTCACGCGAGCCCAATCCCATCCGGACAGTGTGCACCTCAAG	351
DB	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
QY	352	ACTAAGAAAGCCTGTGCATCACCTTGACCTGGGAGACCTTCTCTGGGAGCTGGCTG	411
DB	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu	100
QY	412	CGCGCTGGCCTACTCTCGGAAGTTTCATGGCAGCAAGTGTCTCAACTCTGGGATAGTGC	471
DB	101	AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
QY	472	GACTCTCTCAGGTACTGTGCATCAACCCCTCTAATGCTGTGTATGGCGGTGTGCACACTG	531

Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGAGGAGCGAGATCGGTGTCTTCCCTCTACGACCAAACTTCATCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAGAGTCTGGCACCCTGTGTGCCAAGACGACTGGAAACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GGGCGGGCGCTGCGAGGACATGGGCTATAAGATAATTTTACTCTAGCAAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACAGCTTGTATGAACACAGAGTCCGCGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAGTGTACACAGATGATCGCTTCTTCAAAAGCAGTGTCTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGATAGCTCGGGGTCAACTGAATCAAGCCGACAGACAGATGTGGCGGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGly 260
QY 892 AGCGGCTCCGGGGCTGCGCTGCGAGTGCAGCTGCAGTCCAGACGAGTCCAGTGC 951
Db 261 SerAlaLeuProGlyValArgProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGAGTGCAGTGCAGCGCGCCACTGCTGGTGAAGAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTACATCATGATGCGCATGCGCATTTGGGGATTTTGAGACATCTTTCATG 1071
Db 301 ProLeuAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGGATACCAAGTAGAAAAGTGTCTTCTCAAAATTTATGATCTCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGACATGATGCTGCGTATGAGCTGCGAAGCTCTGACCTTCAAGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCCAGTGTCTGCCCAACCCAGGCTATGCTGCAGCCAGACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTGTGAACTGCGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTTGACACACAGATGCAACAGCATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGCCATGATGTGCGGCTTCTGCGAGGAGACGTCGATTTCTGCCAGGTCAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGGGGCTGCTGCTACCTCGAAGACATATCTGCTGCTGATAGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGCTGTGCGAAAGCTTACAGACAGGAGTGTACGGGATGTGATGCTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGATTTATCGAATAATGAGGCGCAGACGCG 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 12

ABU71851

ID ABU71851 standard; protein; 492 AA.

XX AC ABU71851;

XX XX 10-JUN-2003 (first entry)

XX DE Prostate cancer associated protein #67.

XX KW Prostate cancer; vaccine; Gene therapy; cytostatic; fusion protein;
KW immunogen; cancer; prostate specific antigen; PSA;
KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
PSMA.

XX OS Homo sapiens.

XX PN US2002192763-A1.

XX PD 19-DEC-2002.

XX PF 29-JUN-2001; 2001US-00895793.

XX PR 04-OCT-1999; 99US-0157455P.

XX PR 04-OCT-2000; 2000US-00679272.

XX PR 28-MAR-2001; 2001US-00822827.

XX XX {XUJ}/ XU J.

XX PA {DILL}/ DILLON D C.

XX PA {MITC}/ MITCHAM J L.

XX PA {HARL}/ HARLOCKER S L.

XX PA {JIAN}/ JIANG Y.

XX PA {KALO}/ KALOS M D.

XX PA {FANG}/ FANGER G R.

XX PA {RETT}/ RETTER M W.

XX PA {STOL}/ STOLK J A.

XX PA {DAYC}/ DAY C H.

XX PA {VEDV}/ VEDVICK T S.

XX PA {CART}/ CARTER D.

XX PA {LISX}/ LI S X.

XX PA {WANG}/ WANG A.

XX PA {SKBI}/ SKIKI Y A W.

XX PA {HEPL}/ HEPLER W T.

XX PA {HEND}/ HENDERSON R A.

XX PA {HURA}/ HURAL J.

XX PA {MCNE}/ MCNEILL P D.

XX PA {HOUG}/ HOUGHTON R L.

XX PA {DBAS}/ Y DE BASSOLS C V.

XX PA {FOYT}/ FOY T M.

XX XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

XX PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;

XX PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;

XX PI McNeill PD, Houghton RL, Y De Bassols CV, Foy TM;

XX XX NFI; 2001-245062/25.

XX XX Prostate specific protein and its encoding polynucleotide, useful for the

XX PT treatment and diagnosis of prostate cancer.

XX PT

XX PS Example 2; SEQ ID NO 932; 85pp; English.

XX XX

XX CC The invention describes a fusion protein comprising at least one amino

XX CC acid sequence of immunogenic portions of any of the 3 sequences not

XX CC defined in the specification, or sequences having at least 70 or 90 %

XX CC sequence identity to any one of the 35 sequences defined in the USPTO web

XX CC site, which is encoded by any of the 4 nucleotide sequences not defined

XX CC in the specification. The fusion protein, composition and methods are

XX CC useful for diagnosing, preventing and/or treating cancer, particularly

XX CC prostate cancer. The proteins are useful as markers to indicate the

XX CC presence or absence of cancer. This is the amino acid sequence of a

XX CC prostate cancer therapy associated protein. Note: The sequence data for

Prostate specific protein and its encoding polynucleotide, useful for the treatment and diagnosis of prostate cancer.

Example 2; SEQ ID NO 932; 85pp; English.

The invention describes a fusion protein comprising at least one amino acid sequence of immunogenic portions of any of the 3 sequences not defined in the specification, or sequences having at least 70 or 90 % sequence identity to any one of the 35 sequences defined in the USPTO web site, which is encoded by any of the 4 nucleotide sequences not defined in the specification. The fusion protein, composition and methods are useful for diagnosing, preventing and/or treating cancer, particularly prostate cancer. The proteins are useful as markers to indicate the presence or absence of cancer. This is the amino acid sequence of a prostate cancer therapy associated protein. Note: The sequence data for

CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?docID=US20020192763

XX SQ Sequence 492 AA;

Alignment Scores:
Pred. No.: 3 48e-221 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservatives: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 4 Gaps: 0

US-09-323-597C-1 (1-1738) x ABU71851 (1-492)

QY 112 ATGGCTTGAACCTCAGGCTCAGCAGCTATTGGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGTCCCACTCTCTACGAGGTCAT 231
Db 21 GluProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACCGCTCCCGTCCCGCAGTACGCCCCGAGGTCCTGACGCGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCACCCCGCTCTCGACGAGCCCAATCCCATCCCGGACAGTGCACCTCAAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAACACTGTGCATCAGCTTGACCTCGGACCTTCTCTGGAGCTGGCGTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCTACTCGAAGTTCATCGGACGACAGTGTCCCACTCTGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACCTGCATCAACCCCTCAACTGTGTGATCGGTGTCACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGCAGAGATCGGTGTGCTCTTACGACCAAACTTCATCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAGTCTGGACCTGTGTGCCAAGCAGTGGACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTCGCAGGACATGGGCTATAAGATAATTTTACTCTAGCCCAAGAA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACCGCTTTTGAACCTGAACACAGTCCCGCAATGTCGT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACCAAGTGTGCTGTCTTCAAAAGCAGTGTCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGCGGTCATCTGAACTCAAGCGCCAGAGAGATGTGGCGCGGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCGCGGCGGCTGGCCCTGGCAGTGTGACCTGACCTGACAGACGTCACGCTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCTCATCTACCCCGGAGTGGATCGTCAACAGCCGCCCTCCTCGGTGGA 1011

Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTTGGACGCGCATTTGGGGGATTTGAGACAACTTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGGCGGATACCAAGTAGAAAGAGTATCTCTCAATCCAAATATGACTCCAAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTCGCTGTGATGAGCTGCGAGACCTCTGACTTTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACAGTGTCTGCTGCCAACCCAGCATGATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGCCACCCAGGAGAAAGGAGAACCTCAGAAAGTGTGAAACGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTCTTCTCATTGAGACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGCCCATCATCTGTGCGCGCTTCTGCGAGGGAACCTCGATTCTTGCAGGCTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 ATGAGAGGCGCTCTGTGCTACTTCGAGAACAAATATCTGTGCTGTGCTGATAGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TCGGGTCTGCTGCTGCGCAAGCTTACAGACCAGGAGTGTACGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGGAATTTATCGAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492

RESULT 13
ID ABB95420 standard; protein; 492 AA.
XX AC ABB95420;
XX DT 19-JUL-2002 (first entry)
XX DE Human P1000C SEQ ID NO 932.
XX KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
XX KW gene therapy.
XX OS Homo sapiens.
XX PN US2002022248-A1.
XX PD 21-FEB-2002.
XX PP 12-JAN-2001; 2001US-00759143.
XX PR 25-FEB-1997; 97US-00806099.
XX PR 01-AUG-1997; 97US-00904804.
XX PR 25-FEB-1998; 98US-00020956.
XX PR 25-FEB-1998; 98US-00030607.
XX PR 14-JUL-1998; 98US-00115453.
XX PR 23-SEP-1998; 98US-00159812.
XX PR 15-JAN-1999; 99US-00232149.
XX PR 09-APR-1999; 99US-00288946.
XX PR 13-JUL-1999; 99US-00352616.

PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-0043686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 10-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
XX
PA (XUJJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2002-255649/30.
XX
PT New prostate-specific polynucleotides for diagnosing and treating
PT diseases, in particular prostate cancer, and as markers for the
PT progression of cancer.
XX
PS Claim 2; SEQ ID NO 932; 87bp; English.
XX
CC The present invention provides prostate-specific coding sequences and
CC their encoded proteins. These can be used in the diagnosis and treatment
CC of cancers, particularly prostate cancer. The present sequence is a
CC protein described in the invention
XX
SQ Sequence 492 AA;
XX
Alignment Scores:
Pred. No.: 3,486-221 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 5 Gaps: 0
US-09-323-597C-1 (1-1738) x ABB95420 (1-492)
QY 112 ATGGCTTTGAAGTACAGGTGACAGTATGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGTCCCGGTGCGCCAGTACGCCCGAGGCTCTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

QY 292 TCCAAACCCCGTCTGTGACGACGCCCAATCCCATCCGCGGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCCTGACCCCTGGGACCTTCTCTCGTGGGAGCTGCCTG 411
Db 81 ThrTylLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCGCTGCCCTACTCTGGAAGTTTCATGGGACAGCAAGTGTCCCAACTCTGGAGTAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTCTAACTGTGTGATGGCGGTGCACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAAATCGTGTGCTCTACGGACCAAACTTCTCATCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgueTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAAGTCTCGCCACCTGTGTGCCAAGACGACTGGACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTGCGAGGACATGGGCTATAAGATAATTTTACTCTACCGAAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
QY 712 GTGGATCAGACGCGATCCACGACTTTATCAAACTGAACACAAAGTCCGCGCAATGTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACGTACACAGTGTGCTGTCTTCAAAAGCAGGTGTTCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
QY 832 TGTATAGCTGCGGGGTCAACTTGAACCTCAAGCCGCGCAGACGAGTGTGGCGGCGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
QY 892 AGCGCGTCCCGGGGCGCTGCGCTGCGAGCTGACGTGCACTCCAGACGCTCCAGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCACCCCGAGTGTGATGTCGACAGCCGCCACTCCGCTGGAATA 1011
Db 281 CysGlyGlySerIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAAACAATCCATGGCATTTGGACGGCATTTGCGGGGATTTTGACACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATTTACTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTGGCTGTGAAGCTGCAGAGCTCTGACTTTCAACAGCCTA 1191
Db 341 ThrTylAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCAAGTGTCTGCCCCAACCCAGGCTCATGTCGACGACGAGCTCTGCTGG 1251
Db 361 ValTylProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
QY 1252 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAGTGTCTGAACGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGCTTCTCATTTAGACACAGATGCCAAGACAGATGATCTCTATGACACCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCGCCATGATCTGTGCGCGCTTCTGACAGGGAACGTCGATCTTTGCCAGGCTGAC 1431

Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
1432 AGTGGAGGCGCTCTGCTCAGCTTCCAAAGAACAAATATCTGGTGGCTGATAGGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpLeuIleGlyAspThrSer 460
1492 TGGGGTCTGGCTGGTGGCCAAAGCTTACAGACACAGGAGTGTACGGGATGTGATGCTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
1552 ACGGACTGGATTATCGACAAATGAGGCGAGACGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 14

ABRS4532

ID ABR54532 standard; protein; 492 AA.

XX AC ABR54532;

XX DT 28-AUG-2003 (first entry)

XX DE Prostate tumour specific protein SEQ ID 932.

XX KW Cytostatic; gene therapy; prostate-specific protein; PSP; human;

XX KW immune response; prostate cancer.

XX OS Homo sapiens.

XX XX W0200289747-A2.

XX PD 14-NOV-2002.

XX XX 09-MAY-2002; 2002WO-US014753.

XX PR 09-MAY-2001; 2001US-00852911.

XX PR 29-JUN-2001; 2001US-00895814.

XX PR 10-DEC-2001; 2001US-00012896.

XX PA (CORI-) CORIXA CORP.

XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
PI Kalos MD, Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS;
PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;
PI McNeill PD, Houghton RL, Vinals Y De BassolsC, Foy TM, Watanabe Y;
Deng T;
XX WPI; 2003-167130/16.

XX PT New prostate-specific proteins and genes, useful in gene therapy,
PT particularly for stimulating an immune response in a patient, or treating
PT prostate cancer in a patient, as well as for diagnosing prostate cancer
PT in a patient.

XX PS Example 2; Page 642-643; 691pp; English.

XX CC The present invention relates to novel prostate-specific proteins (PSP)
CC and their coding sequences. The PSPs and their coding sequences are
CC useful for stimulating an immune response in a patient, or for treating
CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention

XX SQ Sequence 492 AA;

Alignment Scores:

Pred. No.:	3,48e-221	Length:	492
Score:	2709.00	Matches:	491
Percent Similarity:	99.80%	Conservative:	0
Best Local Similarity:	99.80%	Mismatches:	1
Query Match:	85.22%	Indels:	0
DB:	6	Gaps:	0

US-09-323-597C-1 (1-1738) x ABR54532 (1-492)

QY 112 ATGGCTTTGAATCAGGGTCCACCAGCTATTGGACCTTACTATGAAAACCATGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGGTCCACACCTGTCTAGAGGTGAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCCTCCCGCTCCCGCTACGCGCGAGGGTCTCTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGCTCTGACAGGAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACGTGTGCATCACCTTGACCTCGGGGACCTTCTCTGGGAGCTCGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGGCTGCCCTACTCTGGAAGTTTCATGGGAGCGAGTGTCTCAACTCTGGGATAGAGTC 471
Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGATGGCGTGTACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAAATCGGTGTTCGCTCTACGACCAAACTTCATTCCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAAGTCTCGGCACCTGTGCGCAAGACGACTGGAACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTCGCAGGACATGGGCTATAAGATAATTTTACTCTAGCCAAAGCAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGGGATCCACAGCTTTTATGAACTGAACACAGTGGCGGCAATCTCGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTACACAGTGCCTGTTCTTCAAAAGCAGTGGTTCCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGATAGCTCGCGGGTCAACTTGAACCTCAAGCCCGCAGAGCAGATTTGTGGCGCGCAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGCGCTGGCCCTGGCAGGTTCAGCTCCAGAACCTCCACGCTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGCGAGCTCCATCATCACCCCGAGTGGATCGTGAAGCGCGCCACTGCTGGTGAATAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGTCATGGACGCTTGGGGGATTTTCAGACAACTCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTATTCTTCATCCAAATATGACTCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAATGACATTGGCGCTGATGAAGCTGCGAGAGCCCTCTGACTTTCAACACCTA 1191

Db 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Qy 1192 GTGAACACGAGTGTCTGCCAACCCAGGACATGATGCTGCAGGCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Qy 1252 ATTTCGGGTGGGGCCACCGAGAGAAAGGAGACCTCAGAGTGCCTGAGCGTGC 1311
Db 381 IISerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAla 400
Qy 1312 AAGGTGCTTCTCATGTAGACACAGAGATGCACAGACAGATGCTGTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Qy 1372 ACACGAGCAGATCTGCGCGCTCTCTGCAGGAGACGTCGATTTCTGCAGCGTGC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Qy 1432 AGTGAGGCGCTCTGCTCACTTCAAGAACAAATATCTGCTGCTGATAGGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Qy 1492 TGGGTTCTGCTGTCGCAAGCTTACAGACAGAGTGTACGGGATGTGATCGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Qy 1552 ACGGACTGATTATCGCAAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaaspGly 492

RESULT 15
ADBI4382
ID ADBI4382 standard; protein; 492 AA.
XX AC ADBI4382;
XX DT 18-DEC-2003 (first entry)
XX DE Human prostate specific protein P1000C.
XX KW Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
XX KW cell therapy; vaccine; T-cell epitope;
XX KW class I major histocompatibility complex allele; MHC; prostate cancer;
XX KW tumour; antigen presenting cell.
XX OS Homo sapiens.
XX PN US2003185830-A1.
XX PD 02-OCT-2003.
XX PF 12-NOV-2002; 2002US-00294025.
XX PR 25-FEB-1997; 97US-00806099.
XX PR 01-AUG-1997; 97US-00904804.
XX PR 09-FEB-1998; 98US-00020956.
XX PR 25-FEB-1998; 98US-00030607.
XX PR 14-JUL-1998; 98US-00115453.
XX PR 23-SEP-1998; 98US-00159812.
XX PR 15-JAN-1999; 99US-00232149.
XX PR 09-APR-1999; 99US-00288946.
XX PR 13-JUL-1999; 99US-00352616.
XX PR 12-NOV-1999; 99US-00439313.
XX PR 18-NOV-1999; 99US-00443686.
XX PR 14-JAN-2000; 2000US-00483672.
XX PR 27-MAR-2000; 2000US-00536857.
XX PR 09-MAY-2000; 2000US-00568100.
XX PR 12-MAY-2000; 2000US-00570737.
XX PR 13-JUN-2000; 2000US-00593793.
XX PR 27-JUN-2000; 2000US-00605783.
XX PR 09-AUG-2000; 2000US-00636215.
XX PR 29-AUG-2000; 2000US-00651236.
XX PR 06-SEP-2000; 2000US-00657279.

PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX
PA (CORI-) CORIXA CORP.
XX Xu J, Stolk JA, Kalos MD;
XX WPI; 2003-756193/71.
DR N-PSDB; ADBI4379, ADBI4380, ADBI4381.

XX New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX
XX Example 2; Page; 101pp; English.

XX The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADBI3563 comprising a sequence ADBI4487. The
CC peptides comprise a fragment ADBI3563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADBI3563 is a polypeptide encoded by a human prostate specific
CC cDNA one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding
CC fragment) that specifically binds to the protein or peptide, detecting
CC the presence of a cancer in a patient (comprising contacting a patient
CC sample with a binding agent that binds to the peptides or a polypeptide
CC appearing as ADBI3558, detecting the amount of polypeptide that binds to
CC the agent and comparing the amount of polypeptide to a predetermined cut-
CC off value to determine the presence of cancer), a fusion protein
CC comprising the peptides or proteins, stimulating or expanding T cells
CC specific for a tumour protein comprising contacting T cells with the
CC peptides or the isolated T cell population, treating prostate cancer in a
CC patient comprising administering a composition comprising the peptides,
CC nucleic acids, antibodies or compounds, determining the presence of a
CC cancer in a patient and treating prostate cancer in a patient comprising
CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC from a patient with the peptides or antigen presenting cells that express
CC (the peptides so that the T cells proliferate, and administering the
CC proliferated T cells to the patient. The peptides (or an oligonucleotide
CC that hybridises to nucleic acid encoding them), is used to detect the
CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC antigen-presenting cells expressing the nucleic acid, are used to
CC stimulate or expand T cells specific for a tumour protein. The peptides,
CC nucleic acids, antibodies, fusion proteins, T cell populations or antigen
CC presenting cells are used to stimulate an immune response or treat
CC prostate cancer in a patient. The present sequence is a prostate specific
CC protein of the invention. Note: Except where otherwise indicated, the
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=20030185830.

XX SQ Sequence 492 AA;

Alignment Scores:
Pred. No.: 3,48e-221 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 7 Gaps: 0

US-09-323-597C-1 (1-1738) x ADBI4382 (1-492)

Qy 112 ATGGCTTGAAGCTCAGGCTCAGCAGCTATTGGACCTTACTATGAAACCATGGATAC 171
|||||

Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Qy 172 CAACCGGAACCCCTATCCCGCACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Qy 232 CCGCTCAGTACTACCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProGlnValLeuThrGlnAla 60
Qy 292 TCCAAACCCCGTCTGTCAGCAGCAGCCAAATCCCACTCCGCGACAGTGTGCACCTCAAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Qy 352 ACTAAGAAAGCACTGTGCATCCTTGCACCTCGGGACCTTCCCTCGGGAGCTGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Qy 412 GCGCTGCGCTTACTCTGAGTCTTGGAGTCTTGGCAGCAGTCTCCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Qy 472 GACTCCTCAGTACTGCATCAACCCCTCTAACTGGTGTGATGCGGTGCACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Qy 532 GCGCGGAGGACGAGATCGGTGTTCGCTCTACGAGCCAAACTTCACTCCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
Qy 592 TACTCATCTCAGAGGAAGTCTGTCACCTGTGTGCCAAGACGACTGGAACGAGAATAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Qy 652 GCGCGGCGGCTCAGGACATGGCTATAGATTAATTTTACTTCTAGCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Qy 712 GTGATGACACGGGATCCACGAGCTTTATGAACCTTGAACACAGTCCCGGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Qy 772 ATCTATAAAGTGTACAGTGTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Qy 832 TGTATAGCTTCGGGGTCAACTTCACTCAAGCGCCAGAGCAGTGTGGCGCGGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Qy 892 AGCGCTCCCGGGGCTGCGCTGCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Qy 952 TCGGAGGCTCCATCATCACCCCGAGTGTGTGACCGCCGCTGCGGCACTGCGTGAATA 1011
Db 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Qy 1012 CCTCTTAACATCCATGTCATGTCAGTGTGCGGGGATTTTGGAGACAATCTTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Qy 1072 TTCTATGAGCGCGGATACCAAGTAGAAGATGATTTCTCATCCAAATTTAGTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Qy 1132 ACCAAGAACATGATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Qy 1192 GTGAAAACGAGTGTGTGTCGCAACCCAGGCGATGATGTGAGCGCAGAACAGCTCTGTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380

Qy 1252 ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAGTGTCTGAACGCTGCC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Qy 1312 AAGGTCTCTCTCATTCAGACACAGAGATGCAACAGCAGATATGCTCTATGACCAACCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Qy 1372 ACACAGCCATGATGTGTCGGGCTTCTGCGAGGGGAAAGTGTGATTTCTGCCAGGCTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Qy 1432 AGTGGAGGCGCTCTGCTCACTTCGAAGAACATATCTGCTGGCTGATAGGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIlePheLeuIleGlyAspThrSer 460
Qy 1492 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Qy 1552 ACGGACTGATTTATCGACAAATGAGGGCAGACGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

Search completed: July 6, 2004, 15:03:08
Job time : 108 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: July 6, 2004, 14:58:46 ; Search time 31.5 Seconds
(without alignments)
10614.659 Million cell updates/sec

Title: US-09-323-597C-1
Perfect score: 3179
Sequence: 1 ggcggaggcgagcgagg.....ctggcaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ r2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09323597/runat_06072004_111728_29674/app_query.fasta_1.1927
-DB=pir_78 -QFWT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09323597 @CGN 1.1.46 @runat_06072004_111728_29674 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	660	20.8	1035	1 A43090	enteropeptidase (E
2	647	20.4	1019	1 A56318	enteropeptidase (E
3	634	19.9	1034	1 A53663	enteropeptidase (E
4	579	18.2	1113	2 J03315	low-density lipopr
5	569.5	17.9	638	1 K0HUP	plasma kallikrein
6	567	17.8	417	1 S00845	hepsin (EC 3.4.21.
7	564.5	17.8	855	2 JC7731	membrane-bound arg
8	556	17.5	638	1 K0RTPL	plasma kallikrein
9	553.5	17.4	812	1 P1XS	plasmin (EC 3.4.21
10	538	16.9	638	1 K0MSPL	plasmin (EC 3.4.21
11	533	16.8	625	1 KPHU1	coagulation factor
12	531	16.7	416	1 S33777	hepsin (EC 3.4.21.
13	504.5	15.9	1524	2 T30337	polyprotein - Afri
14	501	15.8	421	1 S11674	acrosin (EC 3.4.21

ALIGNMENTS

RESULT 1

A43090 enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

N;Alternate names: enterokinase
C;Species: Bos primigenius taurus (cattle)

C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 28-Apr-2003
C;Accession: A43090; A48874; A61436

R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91: 7588-7592, 1994

A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease con
A;Reference number: A43090; MUID:94329561; PMID:8052624

A;Status: nucleic acid sequence not shown; translated from GE/EMBL/DBU
A;Molecule type: mRNA

A;Residues: 1-1035 <KIT>

A;Cross-references: GB:U09859; NID:G746410; PIDN:AAB40026.1; PID:G746411
A;Experimental source: small intestine

R;LaValle, E.R.; Rehentulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.
J. Biol. Chem. 268, 23311-23317, 1993

A;Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of
A;Reference number: A48874; MUID:94043122; PMID:8226855

A;Accession: A48874

A;Molecule type: mRNA

A;Residues: 801-1035 <LAV>

A;Cross-references: GB:U09859; NID:G746410; PIDN:AAB40026.1; PID:G746411

A;Note: parts of this sequence, including the amino end of the mature protein, were co
R;Light, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A;Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A;Reference number: A61436; MUID:92189715; PMID:1799406

A;Accession: A61436

A;Molecule type: protein

A;Residues: 801-807, 'Y', 809-827 <LIG>

C;Comment: The mechanism of association with the membrane of the intestinal brush bor
embrane attachment using a signal-anchor sequence.

C;Comment: Conversion from membrane-bound to soluble forms may involve further process
C;Complex: mature enteropeptidase is variously reported to contain two (heavy and light

lfiide linked

acrosin (EC 3.4.21
acrosin (EC 3.4.21
brain-specific ser
plasmin (EC 3.4.21
plasmin (EC 3.4.21
acrosin (EC 3.4.21
serine proteinase
acrosin (EC 3.4.21
acrosin (EC 3.4.21
acrosin (EC 3.4.21
plasmin (EC 3.4.21
plasmin (EC 3.4.21
proctasin (EC 3.4.
plasmin (EC 3.4.21
apoptasin (EC 3.4.21
apoptasin(a) (EC
hepatocyte growth
nuclei protein prec
trypsin (EC 3.4.21
mast cell tryptase
trypsin (EC 3.4.21
trypsin (EC 3.4.21
trypsin (EC 3.4.21
trypsin (EC 3.4.21
trypsin (EC 3.4.21
trypsin (EC 3.4.21

otated below) or with amino-terminal myristoylation of the heavy chain.

C:Genetics:
A:Gene: GDB:PRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light) ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv ducts.

C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding res
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <RCH>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:536-631/Domain: C1r/C1s repeat homology <C1R>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRQ>
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,446,470,503,534,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,871/Active site: His, Asp, Ser #status predicted

Alignment Scores:

Pred. No.:	1-16e-38	Length:	1019
Score:	647.00	Matches:	138
Percent Similarity:	53.25%	Conservative:	67
Best Local Similarity:	35.84%	Mismatches:	154
Query Match:	20.35%	Indels:	26
DB:	1	Gaps:	10

US-09-323-597C-1 (1-1738) x A56318 (1-1019)

```

QY 448 TGCTCCAACTCTGGATAGAGTGGCTCTCAGTACCTGCATCAACCCCTCTACTGG 507
Db 643 CysLysAlaAspHisPheGlnCys---LysAsnGlyGluCysValProLeuValAsnLeu 661
QY 508 TGTGATGCGCTGTACACTGCGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 567
Db 662 CysAspGlyHisLeuHisCysGluAspGlySerAspGluAlaAspCysValArgPhePhe 681
QY 568 GGA-----CCAACTTCATCTCCTCAGGTGTACTCATCTCAG 693
Db 682 AsnGlyThrThrAsnAsnAsnGlyLeuValArgPheArgileGln----- 696
QY 604 AGGAAGTCTTGGCACCTGTGTGCAAGACGACTGGAACGAGAACTACGGCGGGCGGCC 663
Db 697 ---SerileTrpHisThrAlaCysAlaGluAsnTrpThrThrGlnIleSerAsnAspVal 715
QY 664 TGCAGGGACATGGGCTATAAGATAATTTTACTCTAGCAAGGATATAGTGGATGACGC 723
Db 716 CysGlnLeuLeuGlyLeuGlySerGly---AsnSerSerIysProIlePheSerThrAsp 734
QY 724 GGATCCACACGCTTTATGAACCTGAACACACAGTGCAGGATGCGATCTATATAAAA 793
Db 735 GlyGly---ProPheValIysLeuAsnThrAlaPro-----AspGlyHisLeuLeu 750
QY 784 CTGTACACAGTATGCTGTCTTCAAAAGCAGTGTCTTTCTTACGCTGT-----ATA 837
Db 751 LeuThrProSerGlnGlnCysLeuGlnAspSerLeuLeuArgLeuGlnCysAsnHisLys 770
QY 838 GCTTGGGGTCAACTGAACCTACAGCGC---CAGAGCAGGATTTGGGGGGGAGAGC 894
Db 771 SerCysGlyLysLysLeuAlaAlaGlnAspIleThrProLysIleValGlyGlySerAsn 790
QY 895 GCGCTCCCGGGGGCTGGCGCTGCGCTGAGCTCAGCTCCAGAACGTCACGCTGTGC 954
Db 791 AlaLysGluGlyAlaTrpProTrpValValGlyLeuTyrtyrGlyGlyArgLeuLeuCys 810
QY 955 GGAGGGTCCATCATACCCCGGAGTGGATCGTGACAGCGCGCCCACTGGGTGGAAAACCT 1014

```

```

Db 811 GlyAlaSerLeuValSerSerAspTrpLeuValSerAlaAlaHisCysValTyrGlyArg 830
QY 1015 CTTAACAATCCATGGCATTTGCGCGGATTTGCGGGGATTTTGAGACAATTTTCATGTTTC 1074
Db 831 AsnLeuGluProSerIysTrpThrAlaIleLeuGlyLeuHisMetLysSerAsnLeuThr 850
QY 1075 -----TATGGAGCCGGATACCAAGTAGAAAAGTGTCTCTCATCCAAATTATGACTCC 1128
Db 851 SerProGlnThrValProArgLeuIleAspGluIleValIleAsnProHisIysTrpArg 870
QY 1129 AAGACCAAGAACAATGACATTTGGCTGATGAAGCTGCAGAAGCTCTGACTTTCAACGAC 1188
Db 871 ArgArgLysAspAsnAspIleAlaMetMetHisLeuGluPheLysValAsnTyrThrAsp 890
QY 1189 CTAGTGAACCAAGTGTCTGCCCCAACCCAGGCATGTCTGCAGCCAGACAGCTCTGC 1248
Db 891 TyrIleGlnProIleCysLeuProGluGluAsnGlnValPheProGlyArgAsnCys 910
QY 1249 TGGATTTCCGGGTGGGGGGCCACCGAGGAGGAAGGAGGAGGAGGAGGAGGAGGAGG 1308
Db 911 SerIleAlaGlyTrpGlyThrValValTyrGlnGlyThrThrAlaAsnIleLeuGlnGlu 930
QY 1309 GCCAAGTGCTCTTCATTTGAGACACAGATGCAACAGCAGATATGTCTATGACACCTG 1368
Db 931 AlaAspValProLeuLeuSerAsnGluArgCysGlnGlnGlnMetProGluTyrAsn--- 949
QY 1369 ATCACACACGAGCATGATCTGTGCGCGCTTCCTGCAGGGAAACGCTCATCTTCCAGGGT 1428
Db 950 IleThrGluAsnMetIleCysAlaGlyTyrGluGluGlyGlyIleAspSerCysGlnGly 969
QY 1429 GACAGTGGAGGCGCTCTGCTACTTCGAAAGAACAAATATCTGGTGGCTGATAGGGGATACA 1488
Db 970 AspSerGlyGlyProLeuMetCysGlnGluAsnAsnArgTrpPheLeuAlaGlyValThr 989
QY 1489 AGCTGGGGTCTCGCTGCTGCCAAAGCTTACACACAGGAGTGTACCGGATGTGATGCTGA 1548
Db 990 SerPheGlyTyrLysCysAlaLeuProAsnArgProGlyValTyrAlaArgValSerArg 1009
QY 1549 TTCACGAGCTGGATT 1563
Db 1010 PheThrGluTrpIle 1014

RESULT 3
A53663
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N:Alternate names: enterokinase
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C:Accession: A53663
R:Watanabe, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa
J. Biol. Chem. 269, 19976-19982, 1994
A:Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; MUID:94327548; PMID:8051081
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1014 <MAT>
A:Cross-references: GB:ID30799; NID:9505122; PIDN:BAA06459.1; PID:9505123
A:Note: parts of this sequence, including the amino ends of three chains isolated from
C:Comment: The mechanism of association with the membrane of the intestinal brush bor
otated below) or with amino-terminal myristoylation of the heavy chain.
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>

```



```

QY 442 AGCAAGTGTCTCAACTGTGGATAGAGTCGACTCTCTCAGGTACTGATCAACCCCTCT 501
Db 721 SerPheCysGlnAspAsnGluLeuGluCys---AlaAsnHisGluCysValProArgAsp 739
QY 502 AACTCGTGTGTATGCGGTGTACACTGCGCCCGCGGAGGAGACAGAAATCGGTGTTCGC 561
Db 740 LeuTrpCysAspGlyTrpValAspCysSerAspSerSerAspGluTrpGlyCysValThr 759
QY 562 CTCCTAC-----GGACCAAACTTCATCTCAGGTGTACTCATCTCAGAGGAAGTCC 612
Db 760 LeuSerLysAsnGlyAsnSerSerSerLeuLeuThrValHisLysSerAlaLysGlu--- 778
QY 613 TGGCACCTGTGTCCAGACGACTGGAACGAGAACTACCGCGCGCGGCTCGACGAC 672
Db 779 ---HisHisValCysAlaAspGlyTrpArgGluThrLeuSerGlnLeuAlaCysLysGln 797
QY 673 ATGGGCTATAGAAATAATTTTACTCTAGCCAGGAATAGTGGATGACACGGGATCCACC 732
Db 798 MetGlyLeuGlyGluPro-----SerValThrLysLeuLeuProGlyGlnGluGln 815
QY 733 AGCTTTATGAACACTGAACACAGAGTCGCGCAATGTCGAT-----ATCTATAAAAA 783
Db 816 GlnTrpLeuArgLeuTrpProAsnTrpGluAsnLeuAsnGlySerThrLeuGlnGluLeu 835
QY 784 CTGTACACAGTGTATGCTGCTCTTCAAAAGCAGTGTCTTTACCTGTATAGCC--- 840
Db 836 LeuValTrpArgHisSerCysProSerArgSerGluLeuSerLeuLeuCysSerLysGln 855
QY 841 ---TGGCGGTCAACTTGAACCTCAAGCGCGCAGACAGAGTGTGGCGCGGAGAGCGCG 897
Db 856 AspCysGlyArgArgProAlaAlaArgMetAsnLysArgLeuGlyGlyArgThrSer 875
QY 898 CTCGCGGGCGCTGGCCCTGGAGGTGAGCTGACGCTGCACGCTCCAG---AACGTCCACGTGTC 954
Db 876 ArgProGlyArgTrpProTrpGlnCysSerLeuGlnSerLeuProSerGlyHisIleCys 895
QY 955 GAGAGCTCCATCATCACCCCGAGTGTGTCGACAGCGCCGACGCGCGCGGAA----- 1008
Db 896 GlyCysValLeuLeuAlaLysLysTrpValLeuThrValAlaHisCysPheGluGlyArg 915
QY 1009 -----AAACCTCTTAACATCCATCGCAT 1032
Db 916 GluAspAlaAspValTrpLysValPheGlyLeuAsnAsnLeuAspHisPro----- 933
QY 1033 TGGACGCGCATTCGCGGGATTTGAGACATCTTTCATGTTCTATGAGCGCGATACCAA 1092
Db 934 -----SerGlyPheMetGlnThrArgPhe----- 941
QY 1093 GTACAAAAGTGTATTTCTCATCCAAATATGACTCCAAGACCAAGAACCAATGACATTGGC 1152
Db 942 ValLysThrIleLeuLeuHisProArgTyrSerArgAlaValValAspTyrAspIleSer 961
QY 1153 CTGATGAAGTCGAGAGCCTCTGACTTTCAACGACTAGTGAACAGTGTGTCTGCCC 1212
Db 962 ValValGluLeuSerAspAspIleAsnGluThrSerTyrValArgProValCysLeuPro 981
QY 1213 AACCCAGGCATGATGTGACCGACCAACGACTCTGTGTGATTCGCGGTGGGGGCCACC 1272
Db 982 SerProGluCluTrpLeuGluProAspThrTyrCysTrpIleThrGlyTrpGly----- 999
QY 1273 GAGGAGAAAGGAGACCTCAGAAAGTGTGGAACGCTGCCAAGGTGTCTTCATTGAGACA 1332
Db 1000 HisMetGlyAsnLysMetProPheLysLeuGlnGluGlyGluValArgIleLeuProLeu 1019
QY 1333 CAGAGATGCAACAGAGATATGCTATGACAACTGTATCACACCCATGATCTGTGCC 1392
Db 1020 GluGlnCysGlnSer---TyrPheAspMetLysThrIleThrAsnArgMetIleCysAla 1038
QY 1393 GGCTTCTCGAGGGGAACCTCGATTCTTCCAGGGTGACAGTGGAGGGCCTCTGTGTCACT 1452
Db 1039 GlyTyrGluSerGlyThrValAspSerCysMetGlyAspSerGlyGlyProLeuValCys 1058
QY 1453 TCGAAG---AACATATATCTGGTGGCTGATAGGGGATACAGCTGGGGTCTCGCTGT--- 1506

```

```

Db 1059 GluArgProGlyGlyGlnTrpThrLeuPheGlyLeuThrSerTrpGlySerValCysPhe 1078
QY 1507 GCCAAGCTTACAGACAGAGGTGTCGGAATGTGATGTATTCACGGACTGGATTAT 1566
Db 1079 SerLysValLeuGlyProGlyValTyrSerAsnValSerTyrPheValGlyTrpIleGlu 1098
QY 1567 CGACAAATAGGGCAGCAGCGCTAATCCACATGCTCTTCGTCCTTGAGCTGTTTACAAG 1626
Db 1099 ArgGlnIleTyrIleGln-----ThrPheLeuGlnLys 1109
QY 1627 AAACAATGGG 1638
Db 1110 LysSerGlnGly 1113

RESULT 5
QHUP
Plasma kallikrein (BC 3.4.21.34) precursor - human
N;Alternate names: kininogenin; plasma prekallikrein
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C;Accession: A00921; A37939
R;Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A;Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four
A;Reference number: A00921; MUID:86243359; PMID:3521732
A;Accession: A00921
A;Molecule type: mRNA
A;Residues: 1-638 <CHU>
A;Cross-references: GB:M13143; NID:gl90262; PIDN:AAA60153.1; PID:gl90263
R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A;Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
A;Reference number: A37939; MUID:91152016; PMID:1998666
A;Accession: A37939
A;Molecule type: protein
A;Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-142;143,'X',289-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCM>
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex;
C;Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C;Comment: the enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal
inogen and may also play a role in the renin-angiotensin system by converting prorenin
C;Genetics:
A;Gene: GDB:KLK3
A;Cross-references: GDB:127575; OMIM:229000
A;Map position: 4q35-4q35
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; int
F;1-19/Domain: signal sequence #status predicted <SIG> <AP2>
F;20-638/Product: plasma kallikrein #status predicted <WAT>
F;20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F;110-199/Domain: apple repeat <AP1>
F;200-289/Domain: apple repeat <AP2>
F;291-380/Domain: apple repeat <AP3>
F;391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F;391-621/Domain: trypsin homology <TRI>
F;21-104;47-77;51-57;111-194;137-166;141-147;201-284;227-256;231-237;292-375;322-328;3
F;127;308;396;453;494/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;318-347;340-345/Disulfide bonds: #status predicted
F;390-391/Cleavage site: Arg-Ile (coagulation factor XIa) #status predicted
F;434;483;578/Active site: His, Asp, Ser #status predicted

Alignment Scores:
Pred. No.: 4,23e-33 Length: 638
Score: 569.50 Matches: 155
Percent Similarity: 46.40% Conservative: 64
Best Local Similarity: 32.84% Mismatches: 160
Query Match: 17.91% Indels: 93
DB: 1 Gaps: 23

```

```

Qy 1255 TC CGGCTGGCGGCGCCACCGAGGAGAAAGGGAAGACCTCAGAACTGCTGAACGCTGCCAAG 1311
Db 520 ThrGlyTrpGlyPheSerIysGluIysGlyGluGlnGlnAsnIleLeuGlnLysValAsn 539
Qy 1315 GTGCTTCTCATTTAGACACACAGATGCAACAGCAGATATGCTCTATGACAACTTGATCACA 1374
Db 540 IleProLeuValThrAsnGluGluCysGlnLysArgTyr--GlnAspTyrLysIleThr 558
Qy 1375 CGAGCCATGATCTGTCGGCTTCCTCGGAGGAAAGCTGCAATCTTCCAGGCTGACAGT 1434
Db 559 GlnArgMetValCysAlaGlyTyrLysGluGlyGlyLysAspAlaCysLysGlyAspSer 578
Qy 1435 GGAGGGCCCTCTGCTCACTTCGAAGAAACAATATCTGGTGGCTGATAGGGGATACAAGCTGG 1494
Db 579 GlyGlyProLeuValCysLysHisGlnGlyMetTrpArgLeuValGlyIleThrSerTrp 598
Qy 1495 GGTCTTGGCTGTGCCAAAGCTTACAGACACAGAGTGTACGGGATGTGATGTTATTCACG 1554
Db 599 GlyGluGluCysAlaArgGluGlnProGlyValTyrThrLysValAlaGluTyrMet 618
Qy 1555 GACTGCATTATTCGACAAATAGG---GCAGACGGC 1587
Db 619 AspTrpIleLeuGluLysThrGlnSerSerAspGly 630

RESULT 6
S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C:Accession: S00845
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane
A:Reference number: S00845; MUID:88209431; PMID:2835076
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LEA>
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C:Genetics:
A:Gene: GDB:HPN; TMPRSS1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
P:23-45/Domain: transmembrane #status predicted <TMN>
P:163-400/Domain: trypsin homology <R>
F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted

Alignment Scores:
Pred. No.: 6,23e-33 Length: 417
Score: 567.00 Matches: 143
Percent Similarity: 47.59% Conservatives: 64
Best Local Similarity: 32.87% Mismatches: 167
Query Match: 17.84% Indels: 61
DB: 1 Gaps: 11

US-09-323-597C-1 (1-1738) x S00845 (1-417)

Qy 340 TGCACCTCAAGACTAAGAAAGCATGTGCATCATCTTGACCTGGGACCTTCTCGTG 399
Db 12 CysCysSerArgProLysValAla-----AlaLeuThrAlaGlyThrLeuLeuLeu 28
Qy 400 GGAGCTCGCCTCGCGCTGGCCCTACTCTGGAAGTTCATGGGAGCAAGTGCTCCCAACTCT 459
Db 29 LeuThrAlaIleGlyAla-Ala-----SerTrpAlaIleValAlaValLeuLeu 44
Qy 460 GGGATAGATGGGACTCTCTCAGTACTGTCATCAACCCCTCTAACTGGTGTGATGGGTG 519
Db 44 uArgSerAspGlnGluProLeuTyrProValGln----- 55
Qy 520 TCACACTGCCCGGGGGGAGCAGAGATCGGTGTGTTGCGCTCTACGGACCAAACTTC 579

```

Db 56 -----ValSerSerAlaAspAlaArgLeuMetValPheAspLysThrGlu----- 70
QY 580 ATCTTCAGGTGACTATCTCAGAGCAAGTCTGGCACCTCTGTGTGTCAGAGCACTGG 639
Db 71 -----GlyThrTrpArgLeuLeuCysSerArgSerAs 82
QY 640 AAGCAGAACTACCGGGCGGGCTGCAGGACATGGCTTAAAGATATTTTACTCT 699
Db 82 nAlaArgValAlaGly--LeuSerCysGluLwMetGlyPheLeuArgAlaLeuThrHis 101
QY 700 AGCAGAGAAATAGTGTATGACAGCGATCCACAGCTTTATGAACATGAACAGATGCC 759
Db 102 SerGluLeuAspValArgThrAlaGlyAla-----AsnGlyThrSer 115
QY 760 GGCATATCTCATATATAA-----AACTGTACACAGTATGCC----- 801
Db 116 GlyPhePheCysValAspGluGlyArgLeuProHisThrGlnArgLeuLeuGluValIle 135
QY 802 -----TGTTCCTCAAAGCAGTGGTTTCTTACGCTGTATAGCTCGGGGTC 849
Db 136 SerValCysAspCysProArgGlyArgPheLeuAlaAlaIleCysGlnAspCysGly--- 154
QY 850 AACTTGAACCTAAGCCCGCAGAGCAGATGTGGCGGGGAGAGCGCTCCCGGGGCC 909
Db 155 ---ArgArgLysLeuProValAspArgIleValGlyArgAspThrSerLeuGlyArg 173
QY 910 TGCCCTTCGAGCTGACCTGCACGCTCCAGAACTCCAGTGTGCGAGGCTCCATCATC 969
Db 174 TrpProTrpGlnValSerLeuArgTyrAspGlyAlaHisLeuCysGlySerLeuLeu 193
QY 970 ACCCGGAGTGGATCGACAGCGCCACTGGTGGMAAACTCTTAACATCATCGG 1029
Db 194 SerGlyAspTrpValLeuThrAlaAlaHisCysPheProGluArgAsnArgValLeuSer 213
QY 1030 CATTCGACGGCATTTGCGGGATTTTCAGACAACTCTTCATGTTCTAGCGCGGATAC 1089
Db 214 ArgTrpArgValPheAlaGlyAlaValAlaGlnAlaSerPro---HisGlyLeuGlnLeu 232
QY 1090 CAAGTAGAAAGATGATTTCTTCATCCAAATAT-----GACTCCAG 1131
Db 233 GlyValGlnAlaValValTyrHisGlyGlyTyrLeuProPheArgAspProAsnSerGlu 252
QY 1132 ACCAAGAACATCAGATTCGCTGTGATGAAGCTCAGAGCTCTGACTTTCAACGACCTA 1191
Db 253 GluAsnSerAsnAspIleAlaLeuValHisLeuSerSerProLeuProLeuThrGluTyr 272
QY 1192 GTGAACCAAGTGTGTCTGCCAACCCAGGCGATGATCTGCAGCCAGACAGCTCTCTGG 1251
Db 273 IleGlnProValCysLeuProAlaAlaGlyGlnAlaLeuValAspGlyLysIleCysThr 292
QY 1252 ATTTCCGGGGGGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1311
Db 293 ValThrGlyTrpGlyAsnThrGlnTyrTyrGlyGlnAlaGlyValLeuGlnGluAla 312
QY 1312 AAGGTGCTTCTCATTGAGACACAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
Db 313 ArgValProIleIleSerAsnAspValCysAsnGlyAlaAspPheTyrGlyAsnGlnIle 332
QY 1372 ACACGACCATGATCTGTGCGGCTCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1431
Db 333 LysProLysMetPheCysAlaGlyTyrProGluGlyGlyIleAspAlaCysGlnGlyAsp 352
QY 1432 AGTGGAGGGGCTCTGTGTC-----ACTTCGAAGAACAAATATCTGGTGGCTGATA 1479
Db 353 SerGlyGlyProPheValCysGluAspSerIleSerArgThrProArgTrpArgLeuCys 372
QY 1480 GGGGATACAGCTGGGGTCTGCTGTGCCAAAGCTTACAGACAGGAGTGAAGGAAAT 1539
Db 373 GlyIleValSerTrpGlyThrGlyCysAlaLeuAlaGlnLysProGlyValTyrThrLys 392
QY 1540 GTGATGGTATTCAGGACTGATTTATCGACAAATGAGG 1578

Db 393 ValSerAspPheArgGluTrpIlePheGlnAlaIleLys 405
RESULT 7
JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
C:Accession: JC7731, JC7775
R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda, J.; Biochem. 130, 425-430, 2001
A:Title: Characterization of a membrane-bound arginine-specific serine proteinase from :
A:Reference number: JC7731; MUID:21421307; PMID:11530019
A:Accession: JC7731
A:Molecule type: mRNA
A:Residues: 1-855 <KIS>
A:Cross-references: DDBJ:AB049189
A:Experimental source: strain Male, 7-week-old
R:Satom, S.; Yamazaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.; Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A:Title: A role for membrane-type serine proteinase (MT-Sp1) in intestinal epithelial ti
A:Reference number: JC7775; PMID:11573963
A:Contents: Small intestine
A:Accession: JC7775
A:Molecule type: mRNA
A:Residues: 1-855 <SAT>
A:Cross-references: DDBJ:AB037898
C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine protea
of specific proteins or peptides on the brushborder membranes. It also participates
lial migration and/or cell loss.
C:Genetics:
A:Gene: MT-Sp1
A:Map position: basolateral cell surface
C:Superfamily: membrane-bound arginine-specific serine proteinase
C:Keywords: protein digestion
Alignment Scores:
Pred. No.: 9,848-33 Length: 855
Score: 564.50 Matches: 128
Percent Similarity: 48.9% Conservatives: 62
Best Local Similarity: 32.9% Mismatches: 131
Query Match: 17.7% Indels: 67
DB: 2 Gaps: 11
US-09-323-597C-1 (1-1738) x JC7731 (1-855)
QY 439 GCACCAAGTGTCTCCAACTCTGGGATAGAGTGGGAGTCTCTAGGTACCTGATCAACCCC 498
Db 522 GlyCysSerCysProAlaGlySerPheLysCys---SerAsnGlyLysCysLeuProGln 540
QY 499 TCTAACTGT 558
Db 541 SerGlnGlnCysAsnGlyLysAspAspCysGlyAspGlySerAspGluAlaSerCysAsp 560
QY 559 CGCCTCTACGGACCAAACTTCTTCAGGTGTACTCATCTCAGGAGGAGTCTCTGGCAC 618
Db 561 AsnValAsnAlaValSerCysThrLysTyrThrTyrArgCysGln-----Asn 576
QY 619 CTTGTGTGTCAGAGCACTGGGAC---CAGAACTACGGCGGGCGGCTTCAGGACATG 675
Db 577 GlyLeuCysLeuAsnLysGlyAsnProGluCysAspGlyLysLysAspCysSerAspGly 596
QY 676 GCGTATAAATAATATTTTACTCTAGCAAGGAATAGTGGATGACAGCGGATTCACACG 735
Db 597 SerAspGluLysAsnCys-----AspCysGlyLeuArgSer 608
QY 736 TTTATGAACATGACACAAAGTCCGGCAATGTCTGATATCTATAAAACTGTACACACT 795
Db 609 PheThrLys----- 611
QY 796 CATGCTGTCTCTCAAAGCAGTGTGTTCTTTACGCTGTATAGCTCGGGGTCAACTTG 855
Db 611 ----- 611

Qy	132	ACCA	CAGCATTATGGAC	TCTTACTATAAAACCATG	GATACCAACCGGAAAAA	CCCCCATATCC	191	
Dd	138	ThrSer	Ser-----ThrThre-	:::				
Qy	192	CGCAC	AGCCCACTGTGTC	CGGCATCTCTAC	GAGTGATCCGGCTCAGT	ACTACCCGTC	251	
Dd	400	SerTrpAla	alaMetPheProHis-					
Qy	252	CCCGT	GCCCAGTACGCC	CCCGAGGT- - - C	CTGACGAGGCTTCCA	ACCCCGTCGTCTG	308	
Dd	408	--ArgHis	serIysThrProGlu	aenPheProAsp	alaGlyLeuGlu	MetAsn--TyrCy	426	
Qy	309	CAGC	AGCCCAATCCCAT	CCGGACAGTGTGC	ACCTCAAGACTAGA	AAGCACTGTG	368	
Dd	426	sArg	aenProAspGly	AspIysGlyPro	TrpCysTyrThr	ThrAsp- - - - -	441	
Qy	369	CATC	ACCTTCACCTGG	GGACCTTCCCTCGT	GGAGCTGGCTGG	CGCTACTCTG	428	
Dd	442	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	446	
Qy	429	GAGT	TTCATGGGAGCA	AG- - - TCGTCCA	ACTCTGGANAGAT	GCGACTCTC	485	
Dd	446	pGlu	TyrCysAsnLeu	IysArgCysSer	GluThrGly- - - - -	- - - - -Glyse	460	
Qy	486	CTGC	ATC- - - - -	- - - - -AACCC	CTTAACTTGGTGT	GATGGCTGCACA	524	
Dd	460	rVal	ValGluLeuPro	ThrValserGln	GluproSerGly	ProSerAspSerGlu	ThrAs	480
Qy	525	CTGC	- - - - -	- - - - -CCCGG	GGGAGGACGAGA	ATOGGTGTGTTG	CTTACCGGACCAA	578
Dd	480	pCys	MelTyrGlyAsn	GlyIysAspTrp	rArgGlyVlvsThr	AlaValThrAla	AlaGlyTh	500

RESULT 9
PLMS
plasmin (EC 3.4.21.7) precursor - mouse
N:Contains: angiotatin; plasminogen
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1991 #sequence revision 01-Nov-1996 #text_change 18-Jun-1999
C:Accession: A38514; S48202; S48203
R:Degen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
Genomics 8, 49-61, 1990
A:Title: Characterization of the cDNA coding for mouse plasminogen and localization of
A:Reference number: A38514; MUID:91184812; EMBL:2081600
A:Accession: A38514
A:Molecule type: mRNA
A:Residues: 1-812 <DEG>
A:Cross-references: GB:J04766; NID:G200402; PIDN:AAAS0168.1; PID:G200403
R:Rijmen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
Eur. J. Biochem. 224, 863-871, 1994
A:Title: Characterization of the murine plasma fibrinolytic system.
A:Reference number: S48202; MUID:95010076; EMBL:7523120
A:Accession: S48202
A:Molecule type: protein
A:Residues: 20-25 <LHJ>
A:Accession: S48203
A:Molecule type: protein
A:Residues: 22-27 <LH2>
C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many on
C:Comment: Plasminogen is converted into plasmin by plasminogen activators, both plasmin
mediately after dissociation from the clot. In the presence of the inhibitor, the activa
e inhibitor, the activation involves also removal of the activation peptide.
C:Comment: Streptolysin I (see PIR:KQWSS1) acts on plasminogen to produce angiotatin. T
eutral in treating solid tumors.

[illegible]

```

RESULT 10
KOMSPL
C:Plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus {house mouse}
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A36557
C:Seidman, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Roche
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compar
A:Reference number: A36557; MUID:91090844; PMID:2264928
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: GB:M58588; NID:G200358; PIDN:AAAG3393.1; PID:G200359
A:Note: part of this sequence, including the amino ends of both the heavy and light C
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-130/Product: plasma kallikrein heavy chain #status experimental <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-139/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:331-638/Product: plasma kallikrein light chain #status experimental <LCH>
F:331-621/Domain: trypsin homology <TRY>
F:21-104, 47-77, 51-57, 111-194, 137-166, 141-147, 201-284, 227-256, 231-237, 292-375, 318-347,
F:137, 215, 308, 396, 494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:434, 483, 578/Active site: His, Asp, Ser #status predicted

Alignment Scores:
Pred. No.: 7,78e-31
Score: 538.00
Percent Similarity: 53.87%
Best Local Similarity: 37.10%
Query Match: 16.92%
DB: 1
Indels: 6
Gaps: 6
Matches: 115
Length: 638
Conservative: 52
Mismatch: 117

```

	664	TGCAGGACATGGGCTATAAGATAATTTTACTCTAGCCCAAGAAATAGTCGAATCACAGC	723
Qy		::: ::: ::: ::: :::	
Dd	340	CyLysgLuGlUyCySLysCySerLeuArgLeuSer-----ThrAspPolySer	356
Qy	724	GGATCCACCACGCTTTATCAAACCGAACACACAAGTCGCCGCAATGTCGATATCTATAAAAAA	783
Dd	357	ProThrArgileThrTyrglyMetGlnGlySerSerGlyTyrserLeuArgLeuCyslys	376
Qy	784	CTGTACACACAGTAGTGCTGTTCTTCAAAGACAGCTGGTTCTTTACGCTGTATAGCTGCG	843
Dd	377	LeuValaspserProaspCysThrThrylsle-----	387
Qy	844	GGGGTCAACTTGAACTCAAGCGCGCAGCAGCAGGATTGTGGGGCGCGAGAGCGCGTCCCG	903
Dd	388	-----AsnAlaargileValGlyGlyThrAsnAlaserLeu	399
Qy	904	GGGGCTGGCCCTGGCAGGTTCAGCTTCAGATCCAGAACGTC-----CACGTGTC	954
Dd	400	GlycInTrpProTrpGlnValserLeuGlnValtylsLeuValserGlnThrHisLeuCys	419
Qy	955	GGAGGCTCCCATCATCACCCCCGAGTGGATCGTGCAGCGCGCCACTGCGTGGTCAAAAACCT	1011
Dd	420	GlyGlySerileileGlyA--ggInTrpValleuthralalaHisCysPheAsp---Gly	438
Qy	1015	CTTAACACATCCATGGCATTGACCGGCATTTCGGGGATTTTCAGACCAATCTTTCATGTC	1074
Dd	439	IleProTyProaspvalTrpArgileTyGlyglyIleLeuSerLeuSerGluIleThr	458
Qy	1075	TATGGAGCC---GGATCCAAGTAGAAAAAGTATTTCTCATCCAAATTATGACTCCCAAG	1131
Dd	459	IvsGlnThrProSerArgeillelvaGluLeullelleHISGlnGluTyrlvsValiser	478

Db 431 HisCysPhe---TyrGlyValGluSerProLysIleLeuArgValTyrSerGlyIleLeu 449
QY 1057 AGACAACTTTTCATG-----TTCTATGAGCGCGATACCAAGTACAAAA 1101
Db 450 AsnGlnSerGluIleLysGluAspThrSerPheGly-----ValGlnGlu 465
QY 1102 GTGATTTCTCACTCAAAATATGACTCCAGAGCAAGAACATGACATTCGCTGATGAG 1161
Db 466 IleIleIleHisAspGlnTyrLysMetAlaGluSerGlyTyrAspIleAlaLeuLeuLys 485
QY 1162 CTGCAGAACCTCTGACTTTTCAACGACCTAGTGAACACAGTGTGTCTGCCAACCCAGGC 1221
Db 486 LeuGluThrThrValAsnTyrThrAspSerGlnArgProIleCysLeuProSerLysGly 505
QY 1222 ATGATGCTGCACCCAGAACAGCTCTGCTGATTTCCGGTGGGGGCCACCGAGAGAA 1281
Db 506 AspArgAsnValIleTyrThrAspCysTrpValThrGlyTyrGlyTyrArgLysLeuArg 525
QY 1282 GGGAGAGACTCAGAAAGTGTGAACCTGCCAAGGTGCTTCTCATGAGACACAGAGATGC 1341
Db 526 AspLysIleGlnAsnThrLeuGlnLysAlaLysIleProLeuValThrAsnGluGluCys 545
QY 1342 AACACAGATATGCTATGACAACTGATCAGACAGCCATGATCTGTGCGCGTCTCTG 1401
Db 546 GlnLysArgTyr---ArgGlyHisLysIleThrHisLysMetIleCysAlaGlyTyrArg 564
QY 1402 CAGGGAGAGCTGATCTTGCAGAGTGCAGAGTGCAGAGCGCTCTGCTCACTTCGAGAAC 1461
Db 565 GluGlyGlyLysAspAlaCysLysGlyAspSerGlyGlyProLysSerCysLysHisAsn 584
QY 1462 AATATCTGTGCTGATAGGGGATACAAAGCTGGGTCTTCTGGTGTGCCAAAGCTTACAGA 1521
Db 585 GluValTrpHisLeuValGlyIleThrSerTrpGlyGluGlyCysAlaGlnArgGluArg 604
QY 1522 CCAGAGCTACGGGAATGATGATGATTCACGAGCTGATTCAGAAATGAGGCA 1581
Db 605 ProGlyValTyrThrAsnValValGluTyrValAspTrpIleLeuGluLysThrGlnAla 624

RESULT 12

S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R:Parley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A:Reference number: S33777; PMID:93305733; PMID:8318546
A:Accession: S33777
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <PAR>
A:Cross-references: EMBL:X70900; NID:g57928; PIDN:CAA50256.1; PID:g57929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status: predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Dissulfide bonds: #status: predicted
F:202,256,352/Active site: His, Asp, Ser #status: predicted

Alignment Scores:
Pred. No.: 2,416-30 Length: 416
Score: 531.00 Matches: 135
Percent Similarity: 44.64% Conservative: 69
Best Local Similarity: 29.54% Mismatches: 183
Query Match: 16.70% Indels: 70
DB: 1 Gaps: 11

US-09-323-597c-1 (1-1738) x S33777 (1-416)
QY 340 TCACCTCAAGACTAAGAAAGCACTGTGTCATCTGACCTTGAACCTGGGCACTTCTCTGTG 399
Db 11 CysSerArgProLysValAla-----AlaLeuThrValGlyThrLeuLeu--- 26

QY 400 GGAGCTGGCTGGCGCTGCTACTCTGGAAGTTTCATGGCAGCAAGTGTCTCAACTCT 459
Db 27 -----PheLeu----- 28
QY 460 GGGATAGTGGCGACTCTCAGGTACCTGCATCAACCCCTCTACTCTGTGTGTGGCGTG 519
Db 29 -----ThrGlyIleGlyAlaAspTrpAlaIleValThr 40
QY 520 TCACACTGCCCGCGGGAGGAGAGATCGGTGTGTTCGCTCTCTACGGACCAACTTC 579
Db 41 IleLeuLeuArgSerAspGlnGluProLeuTyrGlnValGlnLeuSerProGlyAspSer 60
QY 580 ATCTTTTCAAGTGTACTCTATCTCAGAGGAAGTCTCGGCACCTGTGTGTGCAAGACACTGG 639
Db 61 ArgLeuLeuValLeuAspLysThrGluGlyThrTrpArgLeuLeuCysSerSerArgSer 80
QY 640 AACGAGAACTACGGCGCGCGCTGCAGGACATGGGTATATAGAATAATTTTTACTCT 699
Db 81 AsnAlaArgValAlaGlyLeuGlyCysGluGluMetGlyPheLeuArgAlaLeuAlaHis 100
QY 700 AGCCAAAGAAATAGTGTGTGTGATGACGGATCCACCGACTTTTATGAAACTGAACACAAAGTCC 759
Db 101 SerGluLeuAspValArgThrAlaGlyAlaAsn-----GlyThrSerGly 115
QY 760 -----GGCAATGTCTGATATCTATAAAACTGTATACCAAGTATGCC 801
Db 116 PhePheCysValAspGluGlyLeuProLeuAlaGlnArgLeuLeuAspValIleSer 135
QY 802 -----TGTTCTTCAAAAGCAGTGTGTCTTTTACGCTGTATAGCTCTGGGGGTCCAC 852
Db 136 ValCysAspCysProArgGlyArgPheLeuThrAlaThrCysGlnAspCysGly----- 153
QY 853 TTGAACTCAAGCCCGCAGACAGATTTGGCGCGCGAGAGCGCGCTCCCGGGCGCTGG 912
Db 154 ArgArgLysLeuProValAspArgIleValGlyGlnAspSerSerLeuGlyArgTrp 173
QY 913 CCCTGGGAGGTGACCTTCACAGCTCCAGAACGTCCACAGTGTCCAGCTGTGCGAGAGCTCCATCATCACC 972
Db 174 ProTrpGlnValSerLeuArgTyrAspGlyThrHisLeuCysGlyGlySerLeuLeuSer 193
QY 973 CCGAGTGGATCTGACAGCGCGCCACTGGTGGGAAACCTCTTAACTATCCATGCCAT 1032
Db 194 GlyAspTrpValLeuThrAlaAlaHisCysPheProGluArgAsnArgValLeuSerArg 213
QY 1033 TGGACGGCATTTGGCGGATTTTGAGACAACTTTTTCATGTCTTCTATGAGCGCGATACCAA 1092
Db 214 TrpArgValPheAlaGlyAlaValAlaArgThrSerPro---HisAlaValGlnLeuGly 232
QY 1093 GTAGAAAAGTGTATTTCTCTCCAAATTAT-----GACTCCAGACC 1134
Db 233 ValGlnAlaValIleTyrHisGlyGlyTyrLeuProPheArgAspProThrIleAspGlu 252
QY 1135 AAGAACAACTGACATTTGGCTGATGAGCTGCAGAGCTCTGACTTTTCAACGACCTAGTG 1194
Db 253 AsnSerAsnAspIleAlaLeuValHisLeuSerSerSerLeuProLeuThrGluTyrIle 272
QY 1195 AAACCACTGTGTGTGCCAACCCAGGATGATGCTGCAGCGCAGAACAGCTCTGCTGATG 1254
Db 273 GlnProValCysLeuProAlaAlaGlyGlnAlaLeuValAspGlyLysValCysThrVal 292
QY 1255 TCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTCTGAAACGCTGCCAAG 1314
Db 293 ThrGlyTrpGlyAsnThrGlnPheTyrGlyGlnGlnAlaValValLeuGlnGluAlaArg 312
QY 1315 GTGCTTCTCATTCAGACACAGATGCAACAGCAGATATGCTATGATGCAACCTGATCACA 1374
Db 313 ValProIleIleSerAsnGluValCysAsnSerProAspPheTyrGlyAsnGlnIleLys 332
QY 1375 CCAGCCATGATCTGTGCGCTTCTCCAGCGGAAACCTGCACTCTGCGGAGGTGACAGT 1434
Db 333 ProLysMetPheCysAlaGlyTyrProGluGlyGlyIleAspAlaCysGlnGlyAspSer 352

R;Vazquez-Levin, M.H.; Reventos, J.; Gordon, J.W.
Eur. J. Biochem. 207, 23-26, 1992
A;Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence of
A;Reference number: S23499; MUID:92331659; PMID:1628652
A;Accession: S23499
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-421 <VAZ>
A;Cross-references: EMBL:M77378
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
R;Keime, S.
submitted to the EMBL Data Library, December 1989
A;Reference number: S12063
A;Accession: S12063
A;Molecule type: DNA
A;Residues: 1-225, 'R', 227-421 <KEI2>
A;Cross-references: EMBL:X54017
R;Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
Hum. Genet. 84, 125-128, 1990
A;Title: Molecular cloning of human proacrosin cDNA.
A;Reference number: A61022; MUID:90128988; PMID:2298447
A;Accession: A61022
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
R;Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
FEBS Lett. 244, 296-300, 1989
A;Title: Primary structure of human proacrosin deduced from its cDNA sequence.
A;Reference number: S03330; MUID:89153568; PMID:2493394
A;Accession: S03330
A;Molecule type: mRNA
A;Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
A;Cross-references: EMBL:Y00970; NID:928325; PIDN:CAA68784.1; PID:928326
C;Genetics:
A;Gene: GDB:ACR
A;Cross-references: GDB:119645; OMIM:102480
A;Map position: 22q13-22qter
A;Introns: 26/2; 94/2; 189/2; 237/3
C;Superfamily: acrosin; trypsin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase; sperm
P;1-19/Domain: signal sequence #status predicted <SIG>
P;20-421/Product: acrosin #status predicted <WAT>
P;43-421/Product: acrosin light chain #status predicted <LCH>
P;43-285/Domain: acrosin heavy chain #status predicted <HCH>
P;302-379/Region: proline-rich
P;22-210/Binding site: carbohydrate (asn) (covalent) #status predicted
P;25-154/Disulfide bonds: #status predicted
P;729-162/Disulfide bonds: #status predicted
P;773-89/Disulfide bonds: #status predicted
P;88,142,240/Active site: His, Asp, Ser #status predicted
P;177-246/Disulfide bonds: #status predicted
P;209-225/Disulfide bonds: #status predicted
P;236-266/Disulfide bonds: #status predicted
Alignment Scores:
Pred. No.: 3,47e-28 Length: 421
Score: 501.00 Matches: 107
Percent Similarity: 56.87% Conservative: 42
Best Local Similarity: 40.84% Mismatches: 87
Query Match: 15.76% Indels: 26
DB: 1 Gaps: 10
US-09-323-597C-1 (1-1738) x S11674 (1-421)
QY 841 TGGCGGGTCACTTGAACCTCAAGCCGCGAGAGC-----AGGANTGGCGCGGAGAGC 894
Db |||||...::: |||||
29 CysGlyLeuArgPheArgGlnAsnProGlnGlyValArgIleValGlyGlyLeuAla 48
QY 895 GCGTCCCGGGCGCTGCGCTGGCGAGCTCAGCTCGACGTC-----CAG 939
Db |||||...::: |||||
49 AlaGlnHisGlyAlaTrpProTrpMetValSerLeuGlnIlePheArgTyrAsnSerHis 68

QY 940 AACGTCACGCTGTGCGAGGCTCCATCATCACCCCGAGTGGATCGTACAGCGCCAC 999
Db |||||...::: |||||
69 ArgTyrHisThrCysGlySerLeuLeuAsnSerArgTrpValLeuThrAlaHis 88
QY 1000 TGC---GTGAAACCTCTTAACATCCATGCGATTCAGCGCATTTGGGGGATTTG 1056
Db |||||...::: |||||
89 CysPheValGlyLys-----AsnAsnValHisAspTrpArgLeuValPheGlyAlaLys 106
QY 1057 AGACAATCTTTCATGTTCTATGGA-----GCCGGATACCAA----- 1092
Db |||||...::: |||||
107 Glu-----IleThrTyrGlyAsnLysProValLysAlaProLeuGlnGluArg 123
QY 1093 ---GTAGAAAAGTGTATCTCATCCAAATTATGACTCCAGCAAGCAACATGACATT 1149
Db |||||...::: |||||
124 TyrValGlyLysIleIleHisGluLysTyrAsnSerAlaThrGluGlyAsnAspIle 143
QY 1150 GCGCTGATGAAGCTGCAGAAAGCCTCTGACTTTTCAACGACCTAGTGAACACAGTGTCTG 1209
Db |||||...::: |||||
144 AlaLeuValGluIleThrProIleSerCysGlyArgPheIleGlyProGlyCysLeu 163
QY 1210 CCCAAC---CCAGGATGATGCTGCAGCCAGCAAGAACAGCTCTGCTGGATTTCCGGTGGGG 1266
Db |||||...::: |||||
164 ProHisPheLysAlaGlyLeuProArgGlySerGlnSerCysTrpValAlaGlyTrpGly 183
QY 1267 GCCACCGAGGAGAAAGG---AAGACTCAGAAGTGTCTGAACGCTGCCAAGTGTCTTCTC 1323
Db |||||...::: |||||
184 TyrIleGluGluLysAlaProArgProSerIleLeuMetGluAlaArgValAspLeu 203
QY 1324 ATTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGTATCACCAGCCATG 1383
Db |||||...::: |||||
204 IleAspLeuAspLeuCysAsnSerThrGlnTrpTyrAsnGlyArgValGlnProThrAsn 223
QY 1384 ATCTGTGCGGGCTTCCTGCAGGGGAAACGTCGATCTTCCAGGGTGACAGTGGGGCCT 1443
Db |||||...::: |||||
224 ValCysAlaGlyTyrProValGlyLysIleAspThrCysGlnGlyAspSerGlyGlyPro 243
QY 1444 CTGGTC-----ACTTCGAAGAACATATCTGTGTGCTGTATAGGGGATACAGCTGGGT 1497
Db |||||...::: |||||
244 LeuMetCysLysAspSerLysGluSerAlaTyrValValGlyIleThrSerTrpGly 263
QY 1498 TCTGGCTGTGCGCAAGCTTACAGCCAGGAGTGTACGGGAATGTGATGTTATTCACGGAC 1557
Db |||||...::: |||||
264 ValGlyCysAlaLeuAlaLysArgProGlyIleTyrThrAlaThrTrpProTyrLeuAsn 283
QY 1558 TGGATT 1563
Db |||||
284 TrpIle 285
RESULT 15
JX0172
acrosin (EC 3.4.21.10) precursor form 3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
C;Accession: JX0172; JX0138
R;Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.
J. Biochem. 109, 828-833, 1991
A;Title: Structure and organization of the mouse acrosin gene.
A;Reference number: JX0172; MUID:92041732; PMID:1939002
A;Accession: JX0172
A;Molecule type: DNA
A;Residues: 1-436 <WAT>
A;Cross-references: GB:S66245; NID:9238706; PIDN:AAB20293.1; PID:9238707
R;Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
J. Biochem. 108, 785-791, 1990
A;Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its
A;Reference number: JX0138; MUID:91185335; PMID:2127931
A;Accession: JX0138
A;Molecule type: mRNA
A;Residues: 4-436 <YAS>
A;Cross-references: GB:D00754; NID:9220322; PIDN:BAA00651.1; PID:9220323
C;Comment: Acrosin is an acrosomal protease that plays an important role in the initi-
e-binding activity.
C;Genetics:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 15:07:22 ; Search time 91 Seconds
(without alignments)
11890.358 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179

Sequence: 1 ggcggagcgagcgagg.....ctggcaaaaaaaaaa 1738

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 2553080

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09323597/runat_06072004_111730_29788/app_query.fasta_1.1927
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -ENDE=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=PCT -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09323597@cgn2_1_13 @runat_06072004_111730_29788
-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	2717	85.5	492	15	US-10-295-027-586	Sequence 586, App
2	2711	85.3	492	14	US-10-205-823-415	Sequence 415, App
3	2709	85.2	492	9	US-09-759-143-932	Sequence 932, App
4	2709	85.2	492	9	US-09-780-669-932	Sequence 932, App
5	2709	85.2	492	9	US-09-822-827-932	Sequence 932, App
6	2709	85.2	492	9	US-09-895-793-932	Sequence 932, App
7	2709	85.2	492	9	US-09-895-814-932	Sequence 932, App
8	2709	85.2	492	13	US-10-012-896-932	Sequence 932, App
9	2709	85.2	492	14	US-10-144-678A-932	Sequence 932, App
10	2709	85.2	492	14	US-10-294-025-932	Sequence 932, App
11	2704	85.1	492	14	US-10-334-038-6	Sequence 6, Appli
12	2701	85.0	492	12	US-09-825-751A-63	Sequence 63, Appli
13	2696	84.8	492	9	US-09-759-143-895	Sequence 895, App
14	2696	84.8	492	9	US-09-780-669-895	Sequence 895, App
15	2696	84.8	492	9	US-09-879-793-14	Sequence 14, Appl
16	2696	84.8	492	9	US-09-822-827-895	Sequence 895, App
17	2696	84.8	492	9	US-09-895-793-895	Sequence 895, App
18	2696	84.8	492	9	US-09-895-814-895	Sequence 895, App
19	2696	84.8	492	10	US-09-776-191-70	Sequence 70, Appl
20	2696	84.8	492	13	US-10-012-896-895	Sequence 895, App
21	2696	84.8	492	14	US-10-144-678A-895	Sequence 895, App
22	2696	84.8	492	14	US-10-294-025-895	Sequence 895, App
23	2696	84.8	492	15	US-10-295-027-588	Sequence 588, App
24	2696	84.8	492	15	US-10-295-027-871	Sequence 871, App
25	2696	84.8	492	15	US-10-158-214A-37	Sequence 37, Appl
26	2696	84.8	492	16	US-10-188-832-204	Sequence 204, Appl
27	2171	68.3	393	9	US-09-759-143-934	Sequence 934, App
28	2171	68.3	393	9	US-09-780-669-934	Sequence 934, App
29	2171	68.3	393	9	US-09-822-827-934	Sequence 934, App
30	2171	68.3	393	9	US-09-895-793-934	Sequence 934, App
31	2171	68.3	393	9	US-09-895-814-934	Sequence 934, App
32	2171	68.3	393	13	US-10-012-896-934	Sequence 934, App
33	2171	68.3	393	14	US-10-144-678A-934	Sequence 934, App
34	2171	68.3	393	14	US-10-294-025-934	Sequence 934, App
35	2126	66.9	384	9	US-09-981-353-23	Sequence 23, Appl
36	1540	48.4	283	9	US-09-988-975A-1	Sequence 1, Appli
37	1263	39.7	229	14	US-10-235-699-1	Sequence 53, Appl
38	1263	39.7	229	10	US-09-898-837A-53	Sequence 897, App
39	1165	36.6	209	9	US-09-759-143-897	Sequence 897, App
40	1165	36.6	209	9	US-09-780-669-897	Sequence 897, App
41	1165	36.6	209	9	US-09-822-827-897	Sequence 897, App
42	1165	36.6	209	9	US-09-895-793-897	Sequence 897, App
43	1165	36.6	209	9	US-09-895-814-897	Sequence 897, App
44	1165	36.6	209	13	US-10-012-896-897	Sequence 897, App
45	1165	36.6	209	14	US-10-144-678A-897	Sequence 897, App

ALIGNMENTS

RESULT 1

US-10-295-027-586

; Sequence 586, Application US/10295027

; Publication No. US2003023250A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Nataasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevizi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Ros Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; PRIORITY FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

```

; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 586
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-586

Alignment Scores:
Pred. No.: 4,28e-207 Length: 492
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.47% Indels: 0
DB: 15 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-295-027-586 (1-492)

QY 112 ATGGCTTTCAACTCAGGGTCCACCCAGCTATTGGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTy-TyrGluAsnHisGlyTyr 20
QY 172 CAACGGAAACCCCTATCCGACAGCCCATCTGGTCCCTCCCACTGTCTACAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CGGGCTCAGTACTACCCGTCCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTGACGACGACCCCAATCCCATCCGACAGTGTGCACTCCAAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAGAAAGCACTGTGCATCACTTGACCTTGGGACCTTCTCTGTGGAGCTCGGTG 411
DB 81 ThrLysLysAlaLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 412 GCGCTGGCTACTCTGGAAGTTTCATGGCCAGCAAGTGTCCAACTCTGGGATAGTGC 471
DB 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTTAACTGGTGTGATGGGTGTCACTGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAAATCGGTGTTCGCTCTACGGACCAAACTTCATCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
QY 592 TACTATCTCAGAGGAGTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAACTAC 651
DB 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
QY 652 GCGCGGGCGGCTCGACGGGACATGGCTATAGATAATTTTACTCTAGCTAGCAAGATA 711

```

```

DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerClnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACAGCTTATGAACTGAACACAGTCCGCGCAATGTCGAT 771
DB 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTACCAAGTGTCTCTTCAAAAGCAGTGGTCTCTTCAAGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCCGGGGTCAACTTGAACTCAAGCCGACAGACAGAGATTGTGGGCGCGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGCGCTGGCCCTGGCAGTCACTGCGACGCTCCAGAACGTCACGCTG 951
DB 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCACCCCGAGTGGATCTGACAGCCGCCCACTGCTGGTGA 1011
DB 281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTACAAATCCATGCGCTGACCGGCTTGGGGGATTTTCAGACATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgLysSerPheMet 320
QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAACTGATTTCTCTCAATTTATGACTCCAA 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGATGCGCTGATGACGCTGACAGAGCTGCTGCTTCAAGACCTTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACCCAGTGTCTGCTGCCCAACCCAGGCTATGCTGACGACGACAGCACTCTGCTGG 1251
DB 361 ValLysProValCysLeuLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
QY 1252 ATTTCCGGGTGGGGGGCCACCGAGGAGAAAGGAGAGCTCAGAGTGTCTGAACTGCTGCC 1311
DB 381 IleSerGlyTyrGlyAlaThrGluGluGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATGTGACACAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGCAGATGATGTGCGGCTTCTGACGGGAAACGTCGATTTCTTGCAGGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTCGAGGCGCTCTGGTCACTTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGGTCTGCTGTGCAAGTTCACAGACCGAGAGTGTACGGGATGTGATGATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGCACTCGATTTATCGCAAAATGAGGGCGACAGGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 2
US-10-205-023-415
; Sequence 415, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian

```


APPLICANT: Kamathkar, Shubhangi
APPLICANT: Wonsey, Angela M.
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Anderson, Dustin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
PRIORITY FILING DATE: 2002-07-25
PRIORITY FILING DATE: 2002-07-25
PRIORITY FILING DATE: 2001-07-25
PRIORITY FILING DATE: 2001-07-25
PRIORITY FILING DATE: 2001-08-22
PRIORITY FILING DATE: 2001-08-22
PRIORITY FILING DATE: 2001-09-25
PRIORITY FILING DATE: 2001-09-25
PRIORITY FILING DATE: 2001-12-12
PRIORITY FILING DATE: 2001-12-12
PRIORITY FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 415
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-10-205-823-415

Alignment Scores:
Pred. No.: 1,28e-206 Length: 492
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 85.28% Indels: 0
DB: 14 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-205-823-415 (1-492)

QY 112 ATGGCTTGAACCTCAGGCTACCAACAGCTATTGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyTyTyGluAsnHisGlyTy 20
QY 172 CAACCGGAAACCCCTATCCCGGACAGCCCACTGTGTCGCCACTGTCTACGAGGTGAT 231
DB 21 GinProGluAsnProTyTyProAlaGlnProThrValProThrValTyGluValHis 40
QY 232 CCGACTCAGTACTACCGCTCCCGCTGCCAGTACGCCCGAGGCTCTGACGACGCT 291
DB 41 ProAlaGlnTyTyTyProSerProValProGlnTyTyTyProAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGCTGTGTCACGACGACCCCAATCCCATCCGGGACAGTGTGACCTCAAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCCTGACCTCGGAGCTCTCTCTGCGGAGCTGCGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCGCTGCTTCTCTGGAAGTTCTATGGGACAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTyPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGGTACTGCTGATCAACCCCTTCACTGCTGTGTGTGATGGCGTGTACACTGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGCGGAGGACGAGATCGGTGTGTTCGCTCTACGACCAAACTTCTCATCTCTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGAAAGTCTGCGACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 651

DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTy 180
QY 652 GCGCGGCGGCTGTCAGGACATGGGCTATPAAGATAATTTTACTCTACCCAAAGGAATA 711
DB 181 GlyArgAlaAlaCysLysAspMetGlyTyTyLysAsnAsnPheTyTySerSerGlnGlyLe 200
QY 712 GTGGATGACAGCGGATCCACCACTTATGAACCTGAACACAAAGTCCGCGCAATGTGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTTACACACAGTGTCTTCTTCAAAAGCAGTGGTTCTTCTTACGC 831
DB 221 IleTyTyLysLysLeuTyHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTGCGGCTCACTTGAACCTGAGCGCGCAGGACGAGTGTGCGGCGGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCGCGGCGCTGCGCTGCGAGTGTGACAGCTGTGACAGCTCCAGACAGCTC 951
DB 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGAGTGTGATGTGACAGCGCGCTGCGTGGGAAAAA 1011
DB 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTCGACGGCATTTTCCGCGCATTTTGTGACAACTCTTTCATG 1071
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGAGCCCGGATACCAAGTAGAAAAAGTGTATTTCTCATCAAAATTATGATCCAA 1131
DB 321 PheTyTyGlyAlaGlyTyTyGlnValGluLysValIleSerHisProAsnTyTyAspSerLys 340
QY 1132 ACCAAGAACATGACATTCGGCTGTGATGAGAGCTGCAGAGAGCTGTCTTCTCAAGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCCAGTGTCTGCCCCAACCCAGGATGATGCTGCAGCGCAGAACAGCTCTCTCTG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGCGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGTAAGCGTGC 1311
DB 381 IleSerGlyTyTyGlyAlaThrGluGluLysGlyTyTySerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTCAGACACAGATGCAACAGCAGATATCTCTATCAACACCTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyTyValTyTyAspAsnLeu 420
QY 1372 ACACCGACCATGATCTGCGCGCTTCTGCGAGGAGAAAGTGTGATCTTCTGCGAGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGTGCTCACTTCGAAGAACAATATCTGTGTGCTGATAGGGGATACAAG 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TCGGGTCTGCTGCTGCCAAAGCTTACAGACAGGAGTGTACCGGAATGTGATGTGATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyTyGlyAsnValMetValPhe 480
QY 1552 ACGGACTGATTTTATCGACAAATGAGGCGACGCGC 1587
DB 481 ThrAspTrpIleTyArgGlnMetArgAlaAspGly 492

RESULT 3

US-09-759-143-932
Sequence 932, Application US/09759143
Patent No. US2002002248A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.

```

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-759-143-932 (1-492)

QY 112 ATGGCTTTGAATCAAGGGTACACAGACTATTGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACGGAAACCCCTATCCCGCAGAGCCACTGTGGTCCCACTGTCTACAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCGTCCTCCCGTCCCGTACGAGTACGCGGAGGCTCTGACGCGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCAAGCAGCCCAATCCCACTCCGGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGAGCTTCTCGTGGAGCTCGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCCTACTCTGGAAGTTTCATGGGCGAAGTGTCTCCAACTCTGGGATAGATGC 471
DB 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACTGTGCATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGAATCGGTGTGTCTCGCTCTACGACCAAACTTCATCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleGluVal 160
QY 592 TACTCATCTCAGAGAAATCTCGGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAspGluAsnTyr 180

```

RESULT 4

```

US-09-780-669-932
; Sequence 932, Application US/09780669
; Patent No. US2002005197A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.

```

```

QY 652 GGGGGGGCGGCTCGAGGACATGGGCTTATAAGATAATTTTACTCTAGCAAGGAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGGATCCACAGCTTTATGAACTTGAACATGACACAAAGTGGCGGAATGTCCAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACTGTATACACAGTGTCCCTGTTCTTCAAAAGCAGTGGTTTCTTACGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGCGGGTCACTTGAACCTCAAGCCCGCAGACAGAGATTGTGGCGGCGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGGCTGGCCCTGCGAGTGTGACCTGACGCTCCAGACCTCCACGCTG 951
DB 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCCATCATCATCCCCCGAGTGTGATCGTACAGCCGCCACTGCTGGGAAAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACATCCATGSCATTTGACGCGCATTTGGCGGGATTTTGAGACATCTTTCATG 1071
DB 301 ProLeuAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATCGAGCGGATACCAAGTAGAAAAAGTGTATTTCTCATCCAAATTTATGATCCCAAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTTGCTGATGACAGTGTGACAGAGCTCTGACTTTTCAGACCTTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACCAAGTGTCTGCCCCAACCCAGGCATGATGTGACGCCAGAACACAGCTCTGCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAGTGTGGAAGCTGCC 1311
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGTCTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACCAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAACCATGATCTGTGCGGCTTCTCGAGGGAAACCTGCGATTCTTGCAGGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGGTCACTTCGAAGAACAAATATCTGTGGCTGTATAGGGGATACAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
QY 1492 TGGGGTTCTGGCTGTGCAAAAGCTTACAGACCCAGGAGTGTACGGGAATGTGATGGTATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACCGACTGGATTTCAGCAATTCAGGGCAGAGCGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

```


; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.534C1
 ; CURRENT APPLICATION NUMBER: US/09/822,827
 ; CURRENT FILING DATE: 2001-03-28
 ; NUMBER OF SEQ ID NOS: 982
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 932
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-822-827-932

Alignment Scores:

Pred. No.: 1,85e-206 Length: 492
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 85.22% Indels: 0
 DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-822-827-932 (1-492)

```

QY 112 ATGCTTTGAACCTCAGGTGACCCAGCTATTGACCTTACTATGAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaileGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCCACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGCTCAGTACTACCGTCCCGTCCCGCCAGTACGCCCGGAGGTCCTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTCAGCAGCCCAATCCCATCCCGGAGGAGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGATCACCCTTGACCTCGGGACCTTCCCTCGGGAGTGCCTG 411
Db 81 ThrLysLysAlaLeuCysileThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCCCTGCGCTTACTCTGAGTTCATGGGAGGAGTCTCCAACTCTGGGATAGATGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyLeuGluCys 120
QY 472 GACTCTCAGGTACTGATCACCCTCTAACTCGTGTGTGATGCGGTGTCCACTGCCCC 531
Db 121 AspSerSerGlyThrCysileAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTGCTTGGCTCTACGACCAAACTTCTCCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheileLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAAGTCTCGCACCTGTGTGCCAAGACCACTGGAAACGAGAACTAC 651
Db 161 TyrSerSerGlnA-glySerSerTyrHisProValCysGlnAspAspTyrPasnGluAsnTyr 180
QY 652 GGGGGGGGGCTCAGGACATGGGTATAGAAATATTTTCTACTTACCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyile 200
QY 712 GTGATGACACGCGGATCCACAGCTTTATGAACCTGAACCAAGTCCCGCAATGTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACCAAGTATGCTGTGTTCTTCAAAAGCAGTGGTTCTTTACGC 831
Db 221 ileTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGCGGGTCAACTGCACTCAAGCGCCAGAGGATGTGGCGGCGGAG 891

```

```

Db 241 CysileAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgileValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCGCTGGCCCTGCAGGTGCAGCTGCACGTCCAGAACGTCACGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGGAGGTCCCATCATCATCCCGCGAGTGCATGCAGCCGCGCCACCTCGGTGGAAAAA 1011
Db 281 CysGlyLysSerileThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTTGGCGGCAATTTGGCGGGATTTTGACACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyileLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGTACCAAGTAGAAAAGTAGATTCTCATCAATATATGACTCCACG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValileSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAATGACATTCGCTGATGAAGCTGCAGAGCCCTCTGACTTTTCAACGACCTA 1191
Db 341 ThrLysAsnAsnAspPheAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACCAAGTGTCTGTGCCCAACCCAGGATGATCTCGAGCCAGCAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
QY 1252 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGAGAACCTCAGAGCTGTGAACGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTCTCTTCTATTGAGACACAGATGCAACAGAGATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuLeuileGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuile 420
QY 1372 ACACCAACCATGATCTGTGCGGCTTCTCGAGGGGAAAGTTCGATCTTCCAGGGGTGAC 1431
Db 421 ThrProAlaMetileCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTGCTGCTACTTCGAAAGAAACAATATCTGTGCTGTAGGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnileTyrTrpLeuileGlyAspThrSer 460
QY 1492 TGGGGTTCGTGCTGCCAAAGCTTACACAGAGTGTACGGGAATCTGATGTTGTTTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACTGGAATTATCGACAAATGAGCGCAGACGGC 1587
Db 481 ThrAspTyrileTyrArgGlnMetArgAlaAspGly 492

```

RESULT 6

US-09-895-793-932
 ; Sequence 932, Application US/09895793
 ; Publication No. US20020192763A1
 ; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.

```

; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-932

Alignment Scores:
Pred. No.: 1.85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-895-793-932 (1-492)

Qy 112 ATGGCTTTGAATCAGGCTCACCAGCTATTGGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaLeuGlnProThrValValProThrValValHis 20
Qy 172 CAACGGGAAACCCCTATCCCGCAGACCCACTGTGTCCTCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProCysProAlaGlnProThrValValProThrValValHis 40
Qy 232 CCGGCTCAGTACTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
Db 41 ProAlaGlnProThrValValProThrValValProThrValValHis 60
Qy 292 TCCAAACCCCGTCTGCAGCAGCCCAATCCCATCCCGTCCCGTCCCGTCCCGTCCCGT 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Qy 352 ACTAAGAAAGCAGTGTGATCAGCTTACCTTGGGACCTTCTCTGGGAGCTCGCGTG 411
Db 81 ThrLysLysAlaLeuCysLeuThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Qy 412 GCGCTGCGCTACTCTGGAAGTTATGGGCGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyLeuGlyCys 120
Qy 472 GACTCTCAGTACTGATCAACCCCTTAACTGGTGTGATGGCGTGTCACTGCCCC 531
Db 121 AspSerSerGlyThrCysLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Qy 532 GCGGGGAGGAGAGATCGGTGTTCTGCTCTACGACCAAACTTCTCAGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuThrGlySerAsnPheLeuGlnVal 160
Qy 592 TACTCATCTCAGAGAGTCTCGCACCTGTGTGCCAAGACGACTGGAACGAGACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
Qy 652 GGGCGGGCGGCTGCGGGACATGGGCTATAGATATATTTTACTCTAGCCAGGATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyThrLysAsnAsnPheThrSerSerGlnGlyLe 200
Qy 712 GTGATGATCAGCGGATCCACAGCTTTATGAACTGAACACAAAGTGGCGGCAATCTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Qy 772 ATCTATAAAAACCTGTACACAGTGTGCTCTTCAAAGCAGTGTCTTCTTTCAGC 831
Db 11eTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Qy 832 TGTATAGCTTCGGGGTCAACTTGAATCAAGCCGACAGCGAGTGTGGGGCGGAG 891
Db 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgLysSerArgIleValGlyGlu 260
Qy 892 AGCGCGCTCCCGGGCCCTGGCCCTGGCAGTGCAGCTGCACCTCCAGAACGTCCACGTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Qy 952 TCGGAGGCTTCATCATCATCCCGGAGTGGATCGTGCAGCCGCCCTCGCTGGGAAAAA 1011
Db 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Qy 1012 CTTCTTAACAATCATTCGACGCGCATTTTCGGGGATTTTGACAAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgLysPheMet 320
Qy 1072 TTCTATGGAGCCGGATACCAAGTAGAAAAAGTGAATTTCTCATCAAAATATGACTCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Qy 1132 ACCAAGAACATCATCATTCGCTGATGAAGCTGCAGAGCTCTGACTTTTCAAGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Qy 1192 GTGAAACCCAGTGTGTCTGCCAACCCAGGCGATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Qy 1252 ATTTCCGGTGGGGGCCACCCAGAGGAAAGGAGAACCTCAGAGTGTCTGAGCGTGC 1311
Db 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Qy 1312 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuLeuLeuThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuLe 420
Qy 1372 ACACAGCCATGATGTGTGCGGCTTCTGCAGGGGAACTCGATTTCTCCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Qy 1432 AGTGGAGGCGCTCTGTGTCATCTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuGlyAspThrSer 460
Qy 1492 TGGGGTTCTCGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGTTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Qy 1552 ACGGACTGGATTTATCGACAAATGAGGGCGAGCGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 7
US-09-895-814-932
; Sequence 932, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugui
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
```



```

; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Mesgher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012.896
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-932

Alignment Scores:
Pred. No.: 1.85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 13 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-012-896-932 (1-492)

Qy 112 ATGCTTGAAGTACAGGTGACACAGCTATTGGACTTACTATGAATACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrGluAsnHisGlyTyr 20

Qy 172 CAACGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaIleProThrValValProThrValTyrGluValHis 40

Qy 232 CCGCTCAGTACTACCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

Qy 292 TCCAAACCCGCTGTCAGCAGCCCAAAATCCCAATCCCGTCCCGTCCCGTCCCGTCCAAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

Qy 352 ACTAAGAAGCACTGTGCATCACCCTTACCTCGGAGACCTTCCTCGTGGAGTGCCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100

Qy 412 GCGCTGCTTACTCTGGAAGTTCATGGCAGCAGTCTCCCACTCTGGATAGTGC 471
Db 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120

Qy 472 GACTCCTCAGGTACTGCATCAACCCCTCTAACTGGTGTGTGATGCGGTGTGCACACTGCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140

Qy 532 GCGGGGAGGACGAGATCGGTGTGTCCTCTACGACCAACTTCATCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160

Qy 592 TACTCATCTCAGAGAGTCTGCGCACCTGTGTGCAAGACGACTCGAAGCAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180

```

```

Qy 652 GGCGGGCGGCGCTGTCAGGACATGGGCTATAAGATAATATTTTACTTCTAGCCAGGAATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200

Qy 712 GTGATGACAGCGGATCCACAGCTTTATGAACACTGAACACAAAGTCCGGCAATGTGCAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAenValAsp 220

Qy 772 ATCTATAAAAACTGTACACAGTGTGCTGCTTTCAAAAGCAGTGGTTTCTTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240

Qy 832 TGTATAGCTTGGCGGCTCAACTTGAACCTCAAGCCGACAGAGAGTGTGGCGGCGAG 891
Db 241 CysIleAlaCysGlyValaLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260

Qy 892 AGCGCGCTCCCGGGGCGCTGCGCCCTGCGCAGTTCAGCTCCAGACAGTCCACGCTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280

Qy 952 TGGGAGGCTCCATCATCATCACCCCGAGTGTGATCGTGCAGCCGCGCCACTGCGTGA 1011
Db 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300

Qy 1012 CCTTTAAACAATCCATGGCATTTGGACGGCATTTGCCGGGATTTTGAGACAATCTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

Qy 1072 TTCTATGGAGCCGGATACCAAGTAGAAAAGTGAATTTCTCATCCAAATATGACTCCAAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340

Qy 1132 ACCAAGAACAATGACATTTGGCTGTGATGAGCTGCAGAGCTCTGACTTTCACGACCTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360

Qy 1192 GTCAAAACCATGTGTCTGCCCAACCCAGCATGCTGCAGCCAGAACAGTCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380

Qy 1252 ATTTCCGGTGGGGGCGCCACCGAGGAGAAAGGAGAACCTCAGAGTGTGAACTGCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400

Qy 1312 AAGTCTCTTCTGATTCAGACACAGATCCCAACAGCAGATATCTCTATGACACCTGATC 1371
Db 401 LysValLeuLeuLeuGlnThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuLe 420

Qy 1372 ACACCGCCATGATCTGTGCGCGCTTCTGCGGGGAAACGTCGATTTTCCCGAGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

Qy 1432 AGTGGAGGCTCTGCTGCTTCTGAGAACATATCTGCTGGCTGATAGGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuLeuGlyAspThrSer 460

Qy 1492 TGGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480

Qy 1552 ACGGACTGGATTATTCGACAAATACGCGGCGAGCGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

```

RESULT 9

```

US-10-144-678A-932
; Sequence 932, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.

```



```
/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Panger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel X.
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A. W.
/ APPLICANT: Repler, William T.
/ APPLICANT: Hural, John
/ APPLICANT: McNeill, Patricia D.
/ APPLICANT: Houghton, Raymond L.
/ APPLICANT: Vinals y de Bassols, Carlota
/ APPLICANT: Foy, Teresa M.
/ APPLICANT: Watanabe, Yoshihiro
/ APPLICANT: Deng, Ta
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
/ FILE REFERENCE: 210121.427C28
/ CURRENT APPLICATION NUMBER: US/10/144,678A
/ CURRENT FILING DATE: 2002-08-12
/ NUMBER OF SEQ ID NOS: 1033
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 932
/ LENGTH: 492
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-144-678A-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 14 Gaps: 0

US-09-323-597c-1 (1-1738) x US-10-144-678A-932 (1-492)

QY 112 ATGGCTTTGAATCAGGGTCACACAGCTATGGACCTTATGAAACCAATGATATC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGGTCCGCCACTGCTTAACAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCGTCGCCCGTCCCGCAGTACGCCCGAGGGTCTCTGACGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCGTGTCTGACGAGCCCAATCCCAATCCGACGAGTGGACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCACTTGACCTGGGGACCTTCTCTGGGGAGCTGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAAGTTCATGGGACGAGTGTCTCAACTCTGGGATAGATGC 471
DB 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyTyrGluCys 120
QY 472 GACTCTCAGGTACCTGCATCAACCCCTCTAATGTTGTGATGGGTGTGCACACTGCCCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGACAGAAATCGTGTGTCCTCTACGGACCAACTTCATCTCAGGTG 591
```

RESULT 10

US-10-294-025-932

; Sequence 932, Application US/10294025

```
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleuGlnVal 160
QY 592 TACTCATCTCAGAGAGAGTCTCTGGCACCTCTGTGCCAAGACGACTGCAAGCAGAACTAC 651
DB 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
QY 652 GGGGGGGGGCTGCGAGGACATGGGCTATAGATATATTTTACTCTAGCCAGCAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACGAGCTTTATGAACTGAAACAAAGTCCGCGCAATCTCGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACCTGTACCAAGTGCCTGCTCTTCTTCAAAAGCGAGTGTCTTTCAGC 831
DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCACCTTGAACCTCAAGCCGCGCAGACGAGGATGTGGCGCGCAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCCCGGGGGCTGCGCTGCGAGGTCAAGCTGCGACGTCAGAACTCCAGACGTC 951
DB 261 SerAlaLeuProGlyAlaTyrProTyrProTyrProTyrProTyrProTyrProTyr 280
QY 952 TGGGAGGCTCCATCATCATCACCCCGAGTGGATCGTGACAGCGCCCTGCTGCGTGA 1011
DB 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGlu 300
QY 1012 CCTTTAAACAATCCATTCGACGCGCATTTGGCGGATTTTGGAGCAATCTTTCATG 1071
DB 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGGATACCAAGTAGTAAAGTGTCTCTCATCCAAATATGATCTCCAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAAACAATGACATTCGCTGATGAAGTCGAGAGGCTCTGACTTTCAACGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACAGTGTGTGCCCCAACCGGAGCATGTGTGCGAGCCAGACACACTCTGCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGGTGGGGGGCCAGAGAGAAAGGAGACCTCAGAAAGTCTGCAAGCTGCC 1311
DB 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCATTTGACACACAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCAGTGTGCGCGGCTCTCTGCGAGGAACTGATTCCTTTCGCGAGGTCAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGCTGCTCATTCCAGAAACAATATCTGGTGGCTGATAGGGGATACAGC 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGTTCTGGCTGCGCAAGCTTACAGACGAGGTGTACGGAAATGTGATGATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGTGATTTTCACAAATAGGGGCGAGACGGC 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
```



```

; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210421.427G29
; CURRENT APPLICATION NUMBER: US/10/294, 025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-294-025-932

Alignment Scores:
Pred. No.: 1,85e-206 Length: 492
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 85.22% Indels: 0
DB: 14 Gaps: 0

US-09-323-597c-1 (1-1738) x US-10-294-025-932 (1-492)

QY 112 ATGGCTTTGAACATCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCAGCCGACCTGTGGTCCGCTCTACGAGGTGCAT 231
DB 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGCTCCCGCTGCCAGTACGCGGCTCTCTGACGAGGCT 291
DB 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCACACCCGCTGTCTGACGAGCCCAATCCCAATCCGAGCAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCATCAGCTTACCTCGGAGACCTCTCGTGGGAGCTGGCTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAAGTTTCATGGCAGCAAGTGTCCAACTCTGGGATAGATGC 471
DB 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCTCAGTACCTGCATCAACCCCTCTAACTGGTGTGATCGGCTGTGCACACTGCC 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCCGCTGTGTGGCTCTACGGACAAACTTCATCTTCAGGTG 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGAAGTCTGGCACCCTGTGTGCCAAGACGACTGCAAGACTAC 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGCGGCGGCTTCAGGAGCATGGGCTATAAGAATAATTTTACTCTAGCCCAAGGAATA 711
DB 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGCGATCCACAGCTTTATGAATCAACACAAAGTCCGCGCAATGTGCAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAACACTGTACCACAGTATGCTGTTCTTCAAAGCAGTGGTTCTTTACGC 831

```

```

DB 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCCCGCAGAGAGATTTGTGGCGCGAG 891
DB 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCTGGCCCTGGCAGGTACCGCTGACGTCACGTCACAGAACCTCCACGTG 951
DB 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATCACCCCGCAGTGGTGTGAGAGCCGCGCCACTCCGCTGCAAAA 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATTCGACCGCATTTGGGGGATTTGGAGCAATCTTTTCATG 1071
DB 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCCGGATACCAAGTACAAAAGTATTTCTCATCAAAATATGACTCCAG 1131
DB 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTCGCTGATGAGCTGCAAGAGCCCTCTGACTTTCAACGACCTTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAAACGAGTGTGTCTGCCCAACCCAGGATGATGCTGCACCCAGACAGCTCTCTCG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGAAAGCTCAGAAAGTGTGAAACGCTGCC 1311
DB 381 IleSerGlyTyrGlyValAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTCATTGAGACAGAGATGCAACAGACAGATATGTCTATGACAACTGTATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACAGGACCATCATCTGTGCCGCTTCTCGAGGGGAAAGCTGATTTCTTGGCAGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCGCTCTCGTCACTTCGAAGAACAAATATCTGTGTGTGTGATAGGGGATACAA 1491
DB 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TCGGTTCTGCTGTGCGCAAGCTTACAGACGAGGAGTGTACGGGAATGTGTATGCTATTC 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGGACTGGATTTATCGACAAATGAGGCGCAGACGCG 1587
DB 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

RESULT 11
US-10-334-038-6
; Sequence 6, Application US/10334038
; Publication No. US20030138865A1
; GENERAL INFORMATION:
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES USEFUL IN
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 07334-135001
; CURRENT APPLICATION NUMBER: US/10/334,038
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US/09/408,628A
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: US 09/164,159
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: US 09/163,759
; PRIOR FILING DATE: 1998-09-30

```

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 492
; TYPE: PRP
; ORGANISM: Homo sapiens
US-10-334-038-6

Alignment Scores:
Pred. No.: 4,62e-206 Length: 492
Score: 2704.00 Matches: 489
Percent Similarity: 99.80% Conservative: 2
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 85.06% Indels: 0
DB: 14 Gaps: 0

US-09-323-597C-1 (1-1738) x US-10-334-038-6 (1-492)

```
QY 112 ATGGCTTTGAATCACTAGGCTACACAGCTATTTGACCTTACTATGAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGGAAACCCCTATCCCGCAGACAGCCCACTGGTCCCACTGTCTACAGAGTGAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCCCTGCCCGTCCCGCAGTACGCCCGAGGGTCTCGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCCACCCCGGCTCTGACGACGACCCCAATCCCATCCGCGGACAGTGTGACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGATCACTACCTTGCCTGGGAGCTTCTCTCGTGGAGCTGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTTACTTGGAGTTTATGGGACGACAGTCTTCAACTCTGGATGAGAGTC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGlnCys 120
QY 472 GACTCTCAGGTACCTGATCAACCCCTTCACTGCTGTGTGATGGGCTGTCACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTGCTTCCCTTACGGACCAAACTTCAFCCTTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlySerAsnPheIleLeuGlnVal 160
QY 592 TACTCATCTCAGAGGAGTCTGGCAGCCCTGTGTGTCAGACAGCTGGAACGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerThrPheProValCysGlnAspSerThrPheAsnGluAsnTyr 180
QY 652 GGGCGGGCGGCTCGCAGGACATGGGCTATAGAATAATTTTACTTCTAGCCAAAGATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACGCGGATCCACAGCTTTATGAACTTGAACATGAAACAGTGGCGGCAATGTCAT 771
Db 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAAGTGTACACAGTGTGCTGCTTCTTCAAAAGCAGTGGTTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTGGGGGTCACTTGAATCAAGCCGACGAGGAGGATGCTGGCGGCGAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGCGCTCCCGGGGCGCTGGCCCTGCGCAGGTGACGCTGCAGCTCCAGAACCTCCACGTG 951
Db 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
```

```
QY 952 TGGGAGGCTCCATCATCATCCCCCGAGTGGATCGTGACAGCCGCCACTCGCTGGAAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGGCATTTGGACGGCATTTGGCGGATTTTGAGACATTTTCATG 1071
Db 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGGAGCGGATACCAAGTAGAAAAAGTATTCTTCATCCAAATATGACTCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACATGACATTTGGCTGATGAAGCTCGAAGAGCTCTGACTTTCAAGACCTTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTCAAAACCGAGTGTCTGCCCAACCCAGGCAATGCTGCAGCCAGAACAGCTCTGTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnCysTyr 380
QY 1252 ATTTCCGGGTGGGGGCCACCCAGAGGAAAGAGGAGACCTCAGAGTGTGTAAGCTGGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTTATGACACAGAGATGCAACAGCAGATATGTCTATGACAACTCTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACCAACCCATGATGTGTGCCGCTTCTTCAGAGGGAGAACCTCGATTTTCCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGCTCACTCGAGAACCAATATCTGGTGGCTGTAGAGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleThrPheIleGlyAspThrSer 460
QY 1492 TGGGGTCTCGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATCGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACGACATGGATTTATCCAAATGAGGGCGAGCGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
```

RESULT 12

US-09-825-751A-63
; Sequence 63, Application US/09825751A
; Publication No. US20030065140A1
; GENERAL INFORMATION:
; APPLICANT: Curagen Corporation
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Heriman, John L.
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/09/825,751A
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 492
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-825-751A-63

```

RESULT 13
US-09-759-143-895
? Sequence 895, Application US/09759143
? Patent No. US2002002248A1
? GENERAL INFORMATION:
? APPLICANT: Xu, Jiangchun
? APPLICANT: Dillon, Davin C.
? APPLICANT: Mitcham, Jennifer L.
? APPLICANT: Harlocker, Susan L.
? APPLICANT: Jiang, Yuqi
? APPLICANT: Henderson, Robert A.
? APPLICANT: Kalos, Michael D.
? APPLICANT: Fanger, Gary R.
? APPLICANT: Retter, Marc W.
? APPLICANT: Stolk, John A.
? APPLICANT: Day, Craig H.
? APPLICANT: Vedvick, Thomas S.
? APPLICANT: Carter, Barrick
? APPLICANT: Li, Samuel
? APPLICANT: Wang, Aijun
? APPLICANT: Skeiky, Rasir A.W.
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
? TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
? FILE REFERENCE: 210121.427C23
? CURRENT APPLICATION NUMBER: US/09/759,143
? CURRENT FILING DATE: 2001-01-12
? NUMBER OF SEQ ID NOS: 934
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 895
? LENGTH: 492
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-759-143-895

```

Alignment Scores:		8e-206	Length:	492
Pred. No.:	Score:	2701.00	Matches:	487
Percent Similarity:	99.80%	Conservative:	4	
Best Local Similarity:	98.98%	Mismatches:	1	
Query Match:	84.96%	Indels:	84	
DB:	12	Gaps:	0	
US-09-323-597C-1 (1-1738) x US-09-825-751A-63 (1-492)				
QY	112	ATGGCTTTGACTCAGGTCACACAGCTATTGGACCTTACTATGAAACCATGATAC	171	
DB	1	MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20	
QY	172	CAACCGGAAACCCCTATCCCGCACAGCCACTGTGTCCTCCACACTCTCTACGAGTGCA	231	
DB	21	GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis	40	
QY	232	CGCGCTCAGTACTACCGCTGCCCGCTGCCCGACAGTACGCCCGGAGGCTCTGACGAGGT	291	
DB	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60	
QY	292	TCCAAACCCCTGCTGTCAGCGAGCCAAATCCCATCCGGACAGTGTGACACTCAAAAG	351	
DB	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80	
QY	352	ACTAAGAAAGACACTGTGCATCACCTTGACCTGGGACCTCTCTCTGGGAGCTGCGCTG	411	
DB	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu	100	
QY	412	GCCTGCTGCTACTCTGGAAGTTCATGGGACAGCAAGTCTCAACTCTGGGATAGAGTGC	471	
DB	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120	
QY	472	GACTCTTCAGTACTCATCAACCCCTCTAACTGGTGTGATCGCGTGTACACTGCCCC	531	
DB	121	AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro	140	
QY	532	GGCGGGAGACAGAAATCGGTGTTCGGCTCTACGGACAAACTCTACCTTCAGGTG	591	
DB	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet	160	
QY	592	TACTCATCTCAGAGGAAGTCTGGACCCCTGTGTGCCAAGACGACTGGACGAGAACTAC	651	
DB	161	TyrSerSerGlnArgLysSerTyrPheIleProValCysGlnAspSerTyrAsnGluAsnTyr	180	
QY	652	GGCGGGCGGCTCGAGGACATGGGCTATAGAAATAATTTTCTCTACCGAGGATA	711	
DB	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200	
QY	712	GTGATGTACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATTCGAT	771	
DB	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220	
QY	772	ATCTATAAAAACTGTACCAGTATGCCCTGTCTTCTTCAAAGAGTGGTTCTTTACGC	831	
DB	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240	
QY	832	TGTATAGCTCGGGGTCAACTTCGACTCAGCGCCAGACGAGTGTGGGGCGGCGAG	891	
DB	241	CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerAlaIleValGlyGlyGlu	260	
QY	892	AGCGGCTCCCGGGGGCTTGGCCCTTGGCAGTCTAGCTGTGACGTCCAGACGTCACGTG	951	
DB	261	SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal	280	
QY	952	TGGGAGGCTCCATCACCACCCCGAGTGGATCTGTGACGCGGCCACTGCTGGTGA	1011	
DB	281	CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys	300	
QY	1012	CCTCTTAAACAATCCATGCATGGACGGCATTTGCGGGCATTTTGTAGACAATCTTTCATG	1071	

Alignment Scores:

Pred. No.: 2e-205 Length: 492
 Score: 2696.00 Matches: 486
 Percent Similarity: 99.80% Conservative: 5
 Best Local Similarity: 98.78% Mismatches: 1
 Query Match: 84.81% Indels: 0
 DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-759-143-895 (1-492)

```

QY 112 ATGGCTTTGAACCTAGGCTACACAGCTATTGACCTTACTATGAAACAGATGATAC 171
DB 1 MetAlaLeuAsnSerGlySerProProAlaileGlyProTyTyTyGluAsnHisGlyTy 20
QY 172 CAACCGGAAACCCCTATCCGCCACAGCCACTGTGGTCCGCCACTGTCTACAGAGTGAT 231
DB 21 GlnProGluAsnProTyProAlaGlnProThrValValProThrValTyGluValHis 40
QY 232 CCGGCTCAGTACTACCGGTCCCGTGGCCCGAGTAGCCCGAGGCTCTCAGCAGGCT 291
DB 41 ProAlaGlnTyTyProSerProValProGlnTyAlaProArgValLeuThrGlnAla 60
QY 292 TCCAAACCCCGTCTGTGACGAGCCCAATCCCATCCGAGACAGTGTGCACCTCAAG 351
DB 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAAGCACTGTGCAFCACCTTGACCCCTGGGAGCCTTCCTCGTGGAGCTCGGTG 411
DB 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCGCTGGCTACTCTGGAGTTCATGGGCACAGTGTCTCAACTCTGGGATAGAGTGC 471
DB 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyLeuGlyCys 120
QY 472 GACTCTCAGGTACCTGTCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
DB 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACAGAAATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 591
DB 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGAGTCTCGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 651
DB 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
QY 652 GCGCGGGCGGCTGTGAGGACATGGGCTATAGAAATAATTTTACTCTAGCCAAAGATTA 711
DB 181 GlyAlaAlaCysArgAspMetGlyTyLysAsnAsnPhetyrSerSerGlnGlyIle 200
QY 712 GTGGATGACAGGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGAT 771
DB 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACTGTACACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 831
DB 221 IleTyTyLysLysLeuTyHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCCGCAGAGCAGGATTTGGGGCGGCGAG 891
DB 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGGCTCCCGGGGCTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 951
DB 261 SerAlaLeuProGlyAlaTrpProTyProGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TCGGAGGCTCCATCATATCCCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1011
DB 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CTTCTTAAATCCATGCGATTGGACGCTTTGGGGATTTGGGGATTTGGAGCAATCTTTTCATG 1071
DB 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

```

```

QY 1072 TTCTATGGAGCGGATACCAAGTAGAAAAAGTGTATTTCTTCATCAAAATTATGACTCCAAG 1131
DB 321 PheTyGlyAlaGlyTyGlnValGlnLysValIleSerHisProAsnTyAspSerLys 340
QY 1132 ACCAAGAACAAATGACATTTGCGTGTGATGAAGCTGCAGAGGCTCTGACTTTCAAGACCTA 1191
DB 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACCCAGTGTGTGCCCCAACCCAGGCTATGCTGCAGCCAGAACACAGCTCTGTCTGG 1251
DB 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTCTGAAAGCTGCCC 1311
DB 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTTCTATTGAGACACAGAGATGCAACAGACAGATGTCTATGACCAACTCTGATC 1371
DB 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyValTyAspAsnLeuIle 420
QY 1372 ACACAGCCCATGATCTGTGCGGCTTCTGTGCGGGAACCTGCGATTCTTGCAGGGTGAC 1431
DB 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGGAGGCTCTGTCTACTTTCGAAACAATAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1491
DB 441 SerGlyGlyProLeuValThrSerAsnAsnAsnIleTrpLeuIleGlyAspHisSer 460
QY 1492 TGGGGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1551
DB 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyGlyAsnValMetValPhe 480
QY 1552 ACGACTGTGATTTATCGACAAATGAGGCGACAGCGC 1587
DB 481 ThrAspTrpIleTyArgGlnMetLysAlaAsnGly 492

```

RESULT 14

US-09-780-669-895

; Sequence 895, Application US/09780669

; Patent No. US20020051977A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedrick, Thomas S.

; APPLICANT: Carter, Barrick

; APPLICANT: Wang, AiJun

; APPLICANT: Li, Samuel

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; APPLICANT: Hural, John

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Houghcon, Raymond L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C24

; CURRENT APPLICATION NUMBER: US/09/780.669

; CURRENT FILING DATE: 2001-02-09

; NUMBER OF SEQ ID NOS: 943

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 895

; LENGTH: 492

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-780-669-895

Alignment Scores:

Pred. No.: 2e-205 Length: 492
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 84.81% Indels: 0
DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-780-669-895 (1-492)

QY 112 ATGGCTTTGAATCAGGCTCACACAGCTATTGGACCTTACTATGAAAAACCATGGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
QY 172 CAACCGAAACCCCTATCCCGACAGCCACCTGCTGCTCCCACTCTACGAGTGCAT 231
Db 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
QY 232 CCGGCTCAGTACTACCGCTCCCGCTGCCCGCTAGCGAGGCTCTGACGAGGCT 291
Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
QY 292 TCAACACCCCTGCTGTCAGCGACCCCAATCCCAATCCCGGACAGTGTGCACCTCAAG 351
Db 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
QY 352 ACTAAGAAACACTGTGCTACCTCACCTCCGCTGGGACCTTCTCGTGGGAGCTGGCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
QY 412 GCGCTGGCTACTCTCGAAGTTCATGGGACGAAAGTCTCCAACTCTGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACCTGATCAACCTCTACTGCTGATGCTGCTGCTGCTGCTGCTGCTG 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
QY 532 GGGGGGAGGACGAGATCGGTGTGCTGCTCTACGGACCAACTTCTCTCAGGTG 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTGAGGAGTCTGACCCCTGTGTGCCAAGACACTGGACGAGACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
QY 652 GGGGGGCGGCTCAGGACATCGGTGTGCTGCTCTACGGACCAACTTCTCTCAGGTG 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGGATGACCGGATCCACGCTTTATGAACTGAACACAACTGACCAAGTCCGGCAATGCTG 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAGCTTACCAAGTATGCTGCTGCTCTCAAAAGCAGGTCTTCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
QY 832 TGATAGCTCGGGGGCTCAACTTGAATCTCAGCCCGCAGACGAGATTTGGGGCGGAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
QY 892 AGCGGCTCCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 951
Db 261 SerAlaLeuProGlyAlaIleProTyrProGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGGGAGGCTCATCATCACCCCGAGTGTGCTGACAGCCCGCCCACTGCTGCTGCTGCTGCTG 1011
Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAAACATCCATGCGATTTGGACCGCATTTGGCGGATTTTGAGACAACTCTTTCATG 1071

Db 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
QY 1072 TTCTATGAGCGGATACCAAGTAGAARAAGTATTTCTCATCCAAATATGACTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAACAATGACATTGCGCTGATGAAGCTGCAAGACCTCTCTCAACACCTTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACCGTGTCTGCTGCCCAACCCAGGCTGATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGCTGGGGGCCCAACCCAGGAGAAAGGAAGACCTCAGAACTGCTGAACCTGCC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGTGTCTCTCATTTGAGACACAGATGCAACACAGATATGTCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCCATCATCTGTGCGGCTTCTGTCAGGGGAACCTCGATTCTTCCAGGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
QY 1432 AGTGAGGCGCTCTGCTCACTTCGAAGAACAAATATCTGCTGCTGATAGGGGATACAGC 1491
Db 441 SerGlyGlyProLeuValThrSerAsnAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
QY 1492 TGGGCTTCTGCTGCTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACCGACTGATTTATCGACAAATGAGGCGAGACGGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetLysAlaAsnGly 492

RESULT 15

US-09-879-792-14
Sequence 14, Application US/09879792
Patent No. US20020061850A1
GENERAL INFORMATION:
APPLICANT: Xiao, Yonghong
APPLICANT: Gedrich, Richard
TITLE OF INVENTION: Regulation of Human Transmembrane Serine
TITLE OF INVENTION: Protease
FILE REFERENCE: 02973.00035
CURRENT APPLICATION NUMBER: US/09/879,792
CURRENT FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/211,224
PRIOR FILING DATE: 2000-06-13
PRIOR APPLICATION NUMBER: US 60/283,353
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: US 60/283,648
PRIOR FILING DATE: 2001-04-16
PRIOR APPLICATION NUMBER: PCT
PRIOR FILING DATE: 2001-06-12
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-879-792-14

Alignment Scores:

Pred. No.: 2e-205 Length: 492
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 84.81% Indels: 0

{Docket No. US20020061850A1 LIO-81-WO}

DB: 9 Gaps: 0

US-09-323-597C-1 (1-1738) x US-09-879-792-14 (1-492)

QY 112 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGATGAAACCACTGATAC 171

Db 1 MetAlaLeuAenSerGlySerProProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20

QY 172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGTCCCACTGTCTACAGGTGCAT 231

Db 21 GlnProGluAenProTyrProAlaGlnProThrValProThrValTyrGluValHis 40

QY 232 CCGGCTAGTACTACCCGTCGCCCGTGCACAGTACGCCCGAGGCTCTCTGACGAGGCT 291

Db 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

QY 292 TCCAAACCCCGTGTGTCGACGAGCCCAAAATCCCATCCCGGACAGTGTGTCACCTCAAAG 351

Db 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

QY 352 ACTAAGAAAGACTGTGCATCACCTTGACCTCCCTGCGGACCTTCTCTGGGAGCTGGCGTG 411

Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100

QY 412 GCCGCTGCGCTACTCTGGAAGTTTCATGGGACGCAAGTCTCCAACTCTGGGATAGATGC 471

Db 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAenSerGlyIleGluCys 120

QY 472 GACTCCTCAGTACTGTCATCAACCTCACTCACTGCTGATGCTGCTGCTGCTGCTGCTGCTG 531

Db 121 AspSerGlyThrCysIleAenProSerAenThrCysAspGlyValSerHisCysPro 140

QY 532 GCGCGGAGGACGAGATCGGTGTGTGCTGCTCTACGGACCAAACTTCTCAGGTG 591

Db 141 GlyGlyGluAenGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnMet 160

QY 592 TACTCATCTCAGAGGAGTCTGCGACCTGTGTGCGCAGACGACGACGACGACGACGACGAC 651

Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAenAspTrpAenGluAenTyr 180

QY 652 GCGCGGCGCGCTGCGAGGACATGGCTATAGAAATAATTTTACTTAGCCCAAGGAATA 711

Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyIle 200

QY 712 GTGATGACAGCGGATCCACAGCTTTATGAACTGAACACAAAGTGGCGGCAATGTGAT 771

Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp 220

QY 772 ATCTATAAAACATGTACACAGTGTGCTGCTCTCAAAAGCAGTGTGTTCTTTACGC 831

Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240

QY 832 TGTATAGCTCGCGGTCAACTTGAATCAAGCCCGCAGACGAGATGTGGCGCGCAG 891

Db 241 CysLeuAlaCysGlyValAenAenAenSerSerArgGlnSerArgIleValGlyGlyGlu 260

QY 892 AGCGGCTCCCGGGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 951

Db 261 SerAlaLeuProGlyAlaIleProTyrProGlnValSerLeuHisValGlnAenValHisVal 280

QY 952 TCGGAGGCTCCATCATCACCCCGGAGTGTGTCAGCGCGCGCTGCTGCTGCTGCTGCTGCT 1011

Db 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300

QY 1012 CCTCTTAACATCATGCGCATGACGCGCATTTGGGGGATTTGAGACATCTTTCATG 1071

Db 301 ProLeuAenAenProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

QY 1072 TTCTATGAGCGCGGATACCAAGTAGAAAAGTGTATTTCTATCCAAATTTATGACTCCAAG 1131

Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAenTyrAspSerLys 340

QY 1132 ACCAAGAACATGACATTTGCGCTGTGATGAAGCTGCGAGAGCTCTGACTTTCAACGACCTA 1191

Search completed: July 6, 2004, 15:23:10
Job time : 117 secs

Db 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAenLeu 360

QY 1192 GTCAAAACCAAGTGTCTGCCCAACCCAGGATGATGCTGCGACGACAGACAGCTCTGCTGG 1251

Db 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380

QY 1252 ATTTCCGGGTGGGGGCCCCAGGAGGAGAAAGGAGACACCTCAGAAAGTGTGAACGCTGCC 1311

Db 381 IleSerGlyTyrPheAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla 400

QY 1312 AAGGTGCTTCTCATTTGAGACACAGATGCAACAGCAGATATGCTATGACAACTGATC 1371

Db 401 LysValLeuLeuIleGluThrGlnAlaGlyCysAenSerArgTyrValTyrAenAenLeuIle 420

QY 1372 ACACGAGCCATGATCTGCGGCTTCTCCAGGGGAAAGTGCATTTCTTCCAGGGTGAC 1431

Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440

QY 1432 AGTGGAGGGCTCTGCTGCTCACTTCGAAAGAAACAATATCTGGTGGCTGATAGGGGATACAAGC 1491

Db 441 SerGlyGlyProLeuValThrSerAenAenAenIleTyrTrpLeuIleGlyAspThrSer 460

QY 1492 TGGGGTCTGCTGCTGCCAAAGCTTACAGACGAGGAGTGTACGGGAATGTGATGCTATTC 1551

Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480

QY 1552 ACCGACTGGATTTATCGACAATGAGGCGACAGCGC 1587

Db 481 ThrAspTrpIleTyrArgGlnMetLysAlaAenGly 492

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 6, 2004, 14:59:51 ; Search time 24 Seconds
(without alignments)
7477.161 Million cell updates/sec

Title: US-09-323-597C-1

Perfect score: 3179

Sequence: 1 ggcggaggcgagcgagg.....ctggcaaaaaaaaaaaaaa 1738

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool_P/US09323597/tunat_06072004_111728_29711/app_query.fasta_1.1927
-DB=Issued Patents AA -QPMF=fastan -SUPPIX=rai -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09323597 -CGNM 1_27 -tunat_06072004_111728_29711 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/6C COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/6D COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2717	85.5	492	3	US-09-342-749-2
2	2717	85.5	492	4	US-09-691-840-2
3	2696	84.8	492	4	US-09-685-166A-895
4	1540	48.4	283	3	US-08-807-151-1
5	1540	48.4	283	4	US-09-478-957-1
6	1165	36.6	209	4	US-09-685-166A-897
7	888	27.9	454	3	US-09-518-046-2
8	866	27.2	159	3	US-09-518-046-24
9	780.5	24.6	455	3	US-09-261-416-2
10	684	21.5	423	4	US-09-656-002-2
11	676.5	21.3	406	4	US-09-851-588-6
12	676.5	21.3	435	3	US-09-008-271A-6

ALIGNMENTS

Alignment Scores:			
Pred. No.:	2,378-232	Length:	492
Score:	2717.00	Matches:	492
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	85.47%	Indels:	0
DB:	3	Gaps:	0

US-09-323-597C-1 (1-1738) x US-09-342-749-2 (1-492)

QY 112 ATGGGTTTCAACTCAGGGTCCACCAGCTATTGCACCTTACTATGAACCATGATAC 171
Db 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20

13	676.5	21.3	437	4	US-09-851-588-8	Sequence 8, Appli
14	660	20.8	798	1	US-08-200-900A-2	Sequence 2, Appli
15	660	20.8	798	5	PCT-US94-09616-2	Sequence 2, Appli
16	575.5	18.1	855	2	US-09-027-337-2	Sequence 2, Appli
17	575.5	18.1	855	4	US-09-644-600-2	Sequence 2, Appli
18	575.5	18.1	955	4	US-09-654-600A-2	Sequence 2, Appli
19	567	17.8	417	4	US-09-820-002-4	Sequence 4, Appli
20	558.5	17.6	376	4	US-09-820-003-2	Sequence 2, Appli
21	556	17.5	638	2	US-08-681-151-3	Sequence 3, Appli
22	553.5	17.4	248	3	US-08-944-483-63	Sequence 63, Appli
23	553.5	17.4	812	1	US-08-248-629A-1	Sequence 1, Appli
24	553.5	17.4	812	1	US-08-451-932-1	Sequence 1, Appli
25	553.5	17.4	812	1	US-08-452-260-1	Sequence 1, Appli
26	553.5	17.4	812	1	US-08-326-785-1	Sequence 1, Appli
27	553.5	17.4	812	2	US-08-612-788-1	Sequence 1, Appli
28	553.5	17.4	812	2	US-08-605-598B-1	Sequence 1, Appli
29	553.5	17.4	812	2	US-08-429-743-1	Sequence 1, Appli
30	553.5	17.4	812	2	US-08-866-735-1	Sequence 1, Appli
31	553.5	17.4	812	3	US-09-066-028-1	Sequence 1, Appli
32	553.5	17.4	812	4	US-09-192-012-3	Sequence 3, Appli
33	553.5	17.4	812	4	US-09-335-325-1	Sequence 1, Appli
34	553.5	17.4	812	4	US-08-991-761A-12	Sequence 12, Appli
35	553.5	17.4	812	5	PCT-US95-05107-1	Sequence 1, Appli
36	548.5	17.3	902	4	US-09-644-600-10	Sequence 10, Appli
37	548.5	17.3	902	4	US-09-654-600A-10	Sequence 10, Appli
38	541	17.0	416	2	US-09-000-845-2	Sequence 2, Appli
39	534	16.8	356	2	US-08-681-151-1	Sequence 1, Appli
40	528.5	16.6	256	2	US-09-027-337-3	Sequence 3, Appli
41	528.5	16.6	256	4	US-09-644-600-3	Sequence 3, Appli
42	528.5	16.6	256	4	US-09-654-600A-3	Sequence 3, Appli
43	527.5	16.6	418	1	US-08-508-448C-25	Sequence 25, Appli
44	527.5	16.6	418	1	US-09-370-838-62	Sequence 62, Appli
45	527.5	16.6	418	4	US-09-370-838-82	Sequence 82, Appli

Db	61	SerAsnProValValCysThrGlnProLysSer2ProSerGlyThrValCysThrSerLys	80
Qy	352	ACTAAGAAAGACACTGTGCATCACCTTGACCTGGGACCTTCTCGTGGAGCTGGCGTG	411
Db	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
Qy	412	GGCGTGGCCCTACTCTGGAAATTTCATGGGCGAGAGTGTCTCAACTCTGGATAGAGTGC	471
Db	101	AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
Qy	472	GACTCTCTACGATACCTCATCAACCCCTCTAACTGGTGTGATGCGGTGTCCACTGCCCC	531
Db	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
Qy	532	GGGGGGAGCAGAGAAATCGGTGTGTTCGGCTCTACGGACCAAACTTCATCTTCAGGTG	591
Db	141	GlyGlyLeuAspGluAsnArgCysValArgLeuTyrglyProAsnPheIleLeuGlnVal	160
Qy	592	TACTCATCTCAGAGGAAGTCTCTGCACCCCTGTGTGCCAAGACACACTGGAAACGAACTAC	651
Db	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr	180
Qy	652	GGGGGGCGCCCTGCAGGGAATGGGCTATAAGATAATTTTCTCTACCTAGCCAAAGAATA	711
Db	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
Qy	712	GTGGATCAGACGGGATCCACCGCTTTATGAACCTGAACACAGTGCCTGGCAATGTCTGAT	771
Db	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
Qy	772	ATCTATAAAAACTGTATCCACAGTATGCTGTTTCTTCAAAAGCAGTGTCTTTCTTACGC	831
Db	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
Qy	832	TGTATAGCTCGGGGTCAACTGAACCTGAAGCCGCCAGACGAGGATTGTGGGGGGGAG	891
Db	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyLeu	260
Qy	892	AGGCGCTCCGGGGGGCTGGCCCTGGCAGTCTAGCCTGCACGCTCCAGAACGCTCCACGTG	951
Db	261	SerAlaLeuProGlyAlaTrpTrpTrpGlnValSerLeuHisValGlnAsnValHisVal	280
Qy	952	TGGCGAGGCTCCATCATCACCCCGAGTGAATGTGTACAGCCGCCACTCGGTGGAAAAA	1011
Db	281	CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys	300
Qy	1012	CCTCTTAACAACTCAATGCATTTGACGCGCATTTTCGGGGATTTTGACACAACTCTTTCATG	1071
Db	301	ProLeuAsnAsnProTrpPheIleTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
Qy	1072	TTCTATGGAGCCGATACCAAGTAGAAAAAGTGTATTTCTCATCCAAATTATGACTCCAAG	1131
Db	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
Qy	1132	ACCAAGAACAAATGACATTGGCTGATGAAGCTGCAGAGGCTCTGACTTTTCAACGCTTA	1191
Db	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
Qy	1192	GTGAACACAGTGTCTGTGCCCAACCCAGGATATGCTGCAGCCAGAACAGCTCTGTGTGG	1251
Db	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
Qy	1252	ATTTCCGGGTGGGGGCCACCGAGAGAAAGGGAAGACCTTCAGAAGTGTGTGAACGTGCC	1311
Db	381	IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
Qy	1312	AAGTGTCTCTCATTTGACACACAGATGCAACAGCAGATGTCTATGACCAACTCATC	1371
Db	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
Qy	1372	ACACGACCATGATCTGTGGCGGTCTCTGCAGGGGAACGTCGATTCTTCCAGGGTGCAC	1431
Db	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlnValAsnValAspSerCysGlnGlyAsp	440

Qy	1432	AGTGGAGGGCTCTGGTCATCTCGAAGACAAATATCTGGTGGCTGATAGGGGG
Db	441	SerGlyGlyProIeuValThrSerLysAsnAsnIleIrrIrrIleuIleGlyAsp
Qy	1492	TGGGGTCTCGCTGTGTCACCAAGCTTACAGACCAGGAGTGTACCGGATATGTGAT
Db	461	TrpGlySerGlyCysAlaLysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMet
Qy	1552	ACGGACTGGGATTTTATCGACAAATGAGGGCGAGACGGC
Db	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly

RESULT 3

```

US-09-685-166A-895
? Sequence 895, Application US/09685166A
? Patent No. 6630305
? GENERAL INFORMATION:
? APPLICANT: Xu, Jiangchun
? APPLICANT: Dillon, Davin C.
? APPLICANT: Mitcham, Jennifer L.
? APPLICANT: Harlocker, Susan L.
? APPLICANT: Jiang, Yuqui
? APPLICANT: Henderson, Robert A.
? APPLICANT: Kalos, Michael D.
? APPLICANT: Fanger, Gary R.
? APPLICANT: Retter, Marc W.
? APPLICANT: Stoik, John A.
? APPLICANT: Day, Craig H.
? APPLICANT: Vedvick, Thomas S.
? APPLICANT: Carter, Barrick
? APPLICANT: Li, Samuel
? APPLICANT: Wang, Aijun
? APPLICANT: Skeiky, Yasir A.W.
? APPLICANT: Hepler, William
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
? TITL OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
? FILE REFERENCE: 210121.427C21
? CURRENT APPLICATION NUMBER: US/09/685.166A
? CURRENT FILING DATE: 2000-10-10
? NUMBER OF SEQ ID NOS: 898
? SOFTWARE: FastSEQ for Windows Version 3.0
? SEQ ID NO 895
? LENGTH: 492
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-09-685-166A-895

```

QY 352 ACTAAGAAAGCACTGTGCATCACCTTGACCCCTGGGAGACCTTCCCTCGTGGAGCTGCCTG 411
Db 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
QY 412 GCCCTGCGCTACTCTGGAAGTTCATGGGACAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
Db 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
QY 472 GACTCCTCAGTACTCTGCATCAACCCCTCACTTAACTGGTGTGTGATGGGTGACACTGCCCC 531
Db 121 AspSerSerGlyThrCysIleAsnProSerAspTrpCysAspGlyValSerHisCysPro 140
QY 532 GCGGGGAGGACGAGATCGGTGTCTGCTCTACGACCAAACTTCATCTTCAGTGT 591
Db 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnMet 160
QY 592 TACTCATCTCAGAGAACTCTGGCACCCCTGTGTGCCAAGACGACTGGAAGAGAACTAC 651
Db 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
QY 652 GCGGGGCGGCTCGCAGGACATGGCTATAGATATTTTACTCTAGCCAGGATA 711
Db 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
QY 712 GTGATGACAGCGGATCCACAGCTTTATGAACTGAACAACTGCAACAACTGCGCAATGTGAT 771
Db 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
QY 772 ATCTATAAAACTGTACACAGTATGCTGTTCTTCAAAAGCAGTGGTTCCTTACGC 831
Db 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
QY 832 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCGCCGACAGCAGGATTTGGCGGGCGAG 891
Db 241 CysLeuAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
QY 892 AGCGGCTCGCGGGGCTCGCCCTCGCAGCTGAGCTGACGTCAGCAGCAGCAGCTCAGTG 951
Db 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
QY 952 TGCGAGGCTCATCATCACCCCGAGTGGATCGTGACAGCGCGCCACTGCTGCGGAAAA 1011
Db 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
QY 1012 CCTCTTAACAATCCATGCGCATGGACGCGCATTTGGGGGATTTTGAGACAACTTTCATG 1071
Db 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuAspGlnSerPheMet 320
QY 1072 TTCTATGAGCGCGATCCAACTAGAAAAGTGTCTCTCATCCAAATATGACTCCCAAG 1131
Db 321 PheTyrGlyAlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLys 340
QY 1132 ACCAAGAAACAATGACATTTGCGCTGATGAGTGCAGAGCCCTCTGACTTTCACAGACTA 1191
Db 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
QY 1192 GTGAACACAGTGTCTGCCCAACCCAGCGATGATGCTGCGAGCCAGAACACCTCTGCTGG 1251
Db 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
QY 1252 ATTTCCGGTGGGGGGCCACCGAGGAGAAAGAGACCTCAGAACTGCTGCAAGCGTGC 1311
Db 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
QY 1312 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCATATGCTATGACAACTGATC 1371
Db 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
QY 1372 ACACGAGCATCATCTGTGCGGGCTTCTGAGGGGAAAGCTGCTGATTTTCAGAGGTGAC 1431
Db 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

QY 1432 ACTGAGGGCTCTGGTCACTTCGAAGAACATATCTGCTGGCTGATAGGGATACAAGC 1491
Db 441 SerGlyGlyProLeuValThrSerAsnAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
QY 1492 TGGGGTTCCTGGCTGTGCCAAAGCTTACAGACCAAGAGTGTACGGGAATGTGATGGTATTC 1551
Db 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
QY 1552 ACCGACTGATTTATCGACAAATGAGGCGACAGCGC 1587
Db 481 ThrAspTrpIleTyrArgGlnMetLysAlaAsnGly 492

RESULT 4
US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO701
; CLONE: 556016
; US-08-807-151-1

Alignment Scores:
Pred. No.: 3,71e-128 Length: 283
Score: 1540.00 Matches: 282
Percent Similarity: 99.65% Conservative: 0
Best Local Similarity: 99.65% Mismatches: 1
Query Match: 48.44% Indels: 0
DB: 3 Gaps: 0

US-09-323-597C-1 (1-1738) x US-08-807-151-1 (1-283)

QY 739 ATGAACCTGAACACAAGTCCGCAATGTCTATATCTATATAAAACTGTACCACTGAT 798
Db 1 MetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAsp 20

799 GCCTGTTCTTCAAAAGCAGTGGTTCTTTACGCTGTATAGCCTGCGGGCTCAACTTGAAC 858
Db 21 AlaCysSerSerLysAlaValValSerLeuArgCysLeuAlaCysGlyValAsnLeuAsn 40
859 TCAAGCCGCGAGAGAGATTGTGGCGGCGAGAGCGCGCTCCCGGGCGCTGGCCCTGG 918
Db 41 SerSerArgGlnSerArgLeuValGlyGlyGlnSerAlaLeuProGlyAlaTrpProTrp 60
919 CAGGTGAGCTGTCAGCTCCAGAACGTCACGCTGTGGGAGGCTCCATCATCACCCCGAG 978
Db 61 GlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerLeuLeuThrProGlu 80
979 TGGATCGTCAGACGCCCGCCACCTGCGTGGAAACCTCTTAAACAATCCATGGGATGGAGC 1038
Db 81 TrpIleValThrAlaAlaHisCysValGlnLysProLeuAsnProTrpHisTrpThr 100
1039 GCATTGGCGGGATTGTGAGCAATCTTTCATGCTCTATGGAGCGGATACCAAGTAGAA 1098
Db 101 AlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGlu 120
1099 AAGTGTATTCTCATCCAAATATTGACTCCAGACCAAGAACAAATGACATTCGCTGATG 1158
Db 121 LysValIleSerHisProAsnTyrAspSerLysThrLysAsnAspIleAlaLeuMet 140
1159 AAGTGTGAGAGCGCTCTGACTTTCACGACCTAGTGAACCACTGTCTGCCACCAACCA 1218
Db 141 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 160
1219 GGCATGATGCTGCAGCCAGACAGCTCTGCTGATTTCCGGGTGGGGCGCCACCGAGGAG 1278
Db 161 GlyMetLeuGlnProGluGlnLeuCysTyrIleSerGlyTyrGlyAlaThrGluGlu 180
1279 AAGGGAAGACCTCAGAGTGTGAAAGCTGCCAGGTCTCTCAATTGAGACACAGAGA 1338
Db 181 LysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuIleGluThrGlnArg 200
1339 TGCACACAGCAGATATCTATGACAACTGATCACAACCGATGATCTGTCGGGCTTC 1398
Db 201 CysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGlyPhe 220
1399 CTGAGGGAGAGCTGATTTCTGCCAGGTGACAGTGGAGGGCTCTGCTCACTTCGAGAG 1458
Db 221 LeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly**LeuValThrSerLys 240
1459 AACAAATATCTGCTGCTGATAGGGATACAGCTGGGTTCTGCTGTGCCAAGCTTAC 1518
Db 241 AsnAsnIleTrpIlePheLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAlaTyr 260
1519 AGACAGGAGTGTACGGGAATGTGATGTTATTCACGAGCTGGATTTATCGAAATGAGG 1578
Db 261 ArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMetArg 280
1579 GCAGACGCC 1587
Db 281 AlaAspGly 283

RESULT 5

US-09-478-957-1
Sequence 1, Application US/09478957
Patent No. 635048
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/478,957
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,151
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4186
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO01
CLONE: 556016
US-09-478-957-1

Alignment Scores:
Pred. No.: 3,71e-128 Length: 283
Score: 1540.00 Matches: 282
Percent Similarity: 99.65% Conservative: 0
Best Local Similarity: 99.65% Mismatches: 1
Query Match: 48.44% Indels: 0
DB: 4 Gaps: 0
US-09-323-597C-1 (1-1738) x US-09-478-957-1 (1-283)

QY 739 ATGAACCTGAACACAGTCCGCGCAATGTCATATCTATAAAACTCTACACAGTGTAT 798
Db 1 MetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAsp 20
QY 799 GCCTGTTCTTCAAAAGCAGTGGTTCTTTACGCTGTATAGCCTGCGGGCTCAACTTGAAC 858
Db 21 AlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGlyValAsnLeuAsn 40
QY 859 TCAAGCCGCGAGAGAGATTGTGGCGGCGAGAGCGCTCCCGGGCGCTGGCCCTGG 918
Db 41 SerSerArgGlnSerArgIleValGlyGlyGlnSerAlaLeuProGlyAlaTrpProTrp 60
QY 919 CAGGTGAGCTGTCAGCTCCAGAACGTCACGCTGTGGGAGGCTCCATCATCACCCCGAG 978
Db 61 GlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerLeuLeuThrProGlu 80
QY 979 TGGATCGTCAGACGCCCGCCACCTGCGTGGAAACCTCTTAAACAATCCATGGGATGGAGC 1038
Db 81 TrpIleValThrAlaAlaHisCysValGlnLysProLeuAsnProTrpHisTrpThr 100
QY 1039 GCATTGGCGGGATTGTGAGCAATCTTTCATGCTCTATGGAGCGGATACCAAGTAGAA 1098
Db 101 AlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGlu 120
QY 1099 AAGTGTATTCTCATCCAAATATTGACTCCAGACCAAGAACAAATGACATTCGCTGATG 1158
Db 121 LysValIleSerHisProAsnTyrAspSerLysThrLysAsnAspIleAlaLeuMet 140
QY 1159 AAGTGTGAGAGCGCTCTGACTTTCACGACCTAGTGAACCACTGTCTGCCACCAACCA 1218
Db 141 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 160
QY 1219 GGCATGATGCTGCAGCCAGACAGCTCTGCTGATTTCCGGGTGGGGCGCCACCGAGGAG 1278

[illegible]

RESULT 6
US-09-685-166A-897
: Sequence 897. Application US/09685166A

Patent No. 5630305
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSTICS OF PROSTATE CANCER

Alignment Scores:		
Pred. No.:	5,58e-95	209
Score:	1165.00	Matches: 208
Percent Similarity:	100.00%	Conservative: 1
Best Local Similarity:	99.52%	Mismatches: 0
Query Match:	36.65%	Indels: 0
DB:	4	Gaps: 0

US-09-323-597C-1 (1-1738) X US-09-685-166A-897 (1-209)

112 ATGGCTTTGAACTCAGGGTCACCCAGCTATTGGACCTTACTATGAAAACCATGGATAC 171

[illegible]

```

RESULT 7
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: In Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
; US-09-518-046-2

```

Alignment Scores:		
Pred. No.:	2,57e-70	Length:
Score:	888.00	Marches:
Percent Similarity:	59.02%	Conservative:
Best Local Similarity:	46.34%	Mismatches:
Query Match:	27.93%	Indels:
		454
		190
		52
		148
		20

```
DB: 3 Gaps: 10
US-09-323-597C-1 (1-1738) x US-09-518-046-2 (1-454)
QY 379 ACCTGGGACCTTCCTCGTGGAGCTGCGCTGCGCTGCTACTCTGGAAGTTTCATG 438
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 SerLeuGlyllelleAlaLeuileLeuAlaLeuAlaileGlyLeuGlylleHisPhe--- 72
QY 439 GCGAGCAAGTGTCCAACTCTGGATAGAGTGGACTCTCTCAGGTACCTGCATCAACCCC 498
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 -----AspCysSerGlyLys---TyrArgCysArgSerSerPheLysCysilleGluLeu 89
QY 499 TCTAAGTGGTGTGATGCGGTGTCACTGCGCGGGGAGACGAGAGATCGTGTGTT 558
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 IleThrArgCysAspGlyValSerAspCysLysAspGlyGluAspGluTyrArgCysVal 109
QY 559 CGCTCTACGACCAAACTTCCTCTCAGGTGTACTCTCAGAGAAAGTCTCTGCGAC 618
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 ArgValGlyGlyGlnAsnAlaValLeuGlnValPheThrAla-----AlaSerTrpLys 127
QY 619 CCTGTGTCCAGACGACTGGAACGAGACTAGCGGGCGGCGCTCGACGGACATGGGC 678
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ThrMetCysSerAspTrpLysGlyHisTyrAlaAsnValAlaCysAlaGlnLeuGly 147
QY 679 TATAAGAAATATTTTCTCTAGCAAGGAATAGTGTGATGACAGCGGATCCACCGCTTT 738
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 PhePro---SerTyrValSerSerAspAsnLeuArgValSerSerLeuGluGlyGlnPhe 166
QY 739 ATGAACCTGAACACAGTGGCGCAATGTC-----GATATCTATAAAACCTGAC 789
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ArgGluGluPheValSerIleAspHisLeuLeuProAspAspLysValThrAlaLeuHis 186
QY 790 CACAGT-----GATGCTGTCTTCAAAAGCAGTGGTCTTCTTACGCTGTATA 837
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 HisSerValTyrValArgGluGlyCysAlaSerGlyHisValThrLeuGlnCysThr 206
QY 838 GCCTGCGGGGTCAACTGAACTCAAGCGCCAGACGAGATGTGGCGCGCAGAGCGCG 897
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 AlaCysGlyHisArgGlyTyrSer-----SerArgIleValGlyGlyAsnMetSer 224
QY 898 CTCGCGGGGCTGCGCTGCGAGTGCCTGCAAGTCCAGACGCTCCAGCTGTGCGGA 957
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 LeuLeuSerGlnTrpTrpTrpGlnAlaSerLeuGlnPheGlnGlyTyrHisLeuCysGly 244
QY 958 GGTCTCCATCAPACCCCGAGTGGATCGTGCAGCGCCCACTCGCTGGGAAACCTCTT 1017
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 GlySerValIleThrProLeuTrpIleThrAlaHisCysVal---TyrAspLeu 263
QY 1018 AACAACTCCATGCGCATTCGACGGCATTTGCGGGATTTTGACACAATCTTTTCATGTTCTAT 1077
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TyrLeuProLysSerTrpThrIleGlnValGlyLeuVal-----SerLeuLeuAspAsn 281
QY 1078 GGAGCC---GGATACCAAGTAGAAAAGTGTATTTCTATCCAAATTTATGACTCCCAAGACC 1134
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 ProAlaProSerHisLeuValGluLysIleValTyrHisSerLysTyrLysProLysArg 301
QY 1135 AAGAACATGATGCTGCTGATGAGCTGCAGAGCTCTCGACTTTCAACGACCTAGTG 1194
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 LeuGlyAsnAspIleAlaLeuMetLysLeuAlaGlyProLeuThrPheAsnGluMetIle 321
QY 1195 AAACCAAGTGTCTGCCCAACCCAGGCATGTGCTGCAGCCAGCAACAGCTCTGCTGGATT 1254
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 GlnProValCysLeuProAsnSerGluGluAsnPheProAspGlyLysValCysTrpThr 341
QY 1255 TCCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTGTGAACGCTGCCAAG 1314
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 SerGlyTrpGlyAlaThrGluAspGlyGlyAspAlaSerProValLeuAsnHisAlaAla 361
QY 1315 GTGCTTCTCATTGACACAGAGATGCACAGCAGATGTCTATGCAACCTGATCACA 1374
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 ValProLeuIleSerAsnLysIleCysAsnHisArgAspValTyrGlyIleIleIleSer 381
QY 1375 CCAGCCATGATCTGTGCGCGCTTCTGCGAGGGGACGTCGATTTCTGCCGCGGTGACAGT 1434
```

```
Db 382 ProSerMetLeuCysAlaGlyTyrLeuThrGlyGlyValAspSerCysGlnGlyAspSer 401
QY 1435 CGAGGCGCTCTCGTCACTTCGGAACAATATCTGCTGCTGATAGGGGATACAGCTGG 1494
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 GlyGlyProLeuValCysGlnGluArgLeuTrpLysLeuValGlyAlaThrSerPhe 421
QY 1495 GGTTCGTGGTGGCCAACTTACAGACCAGAGTGTACGGGAATCTGATGGTATTTCACG 1554
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 GlyIleGlyCysAlaGluValAsnLysProGlyValTyrThrArgValThrSerPheLeu 441
QY 1555 GACTGGATTATCGACAAATGAGGGCGAC 1584
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442 AspTrpIleHisGluGlnMetGluArgAsp 451
RESULT 8
US-09-518-046-24
; Sequence 24, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 24
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; OTHER INFORMATION: protease domain of TMPRSS2 (Tmprss2)
US-09-518-046-24
Alignment Scores:
Pred. No.: 1,56e-68 Length: 159
Score: 866.00 Matches: 158
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.37% Mismatches: 0
Query Match: 27.24% Indels: 0
DB: 3 Gaps: 0
US-09-323-597C-1 (1-1738) x US-09-518-046-24 (1-159)
QY 976 GACTGGATCTGACAGCCGCCCACTCGTGGAAAACTCTTAACAATCCATGCGATTGG 1035
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 GluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnProTrpHisTrp 20
QY 1036 ACGGCATTTTCGGGGGATTTTGAGACAATCTTTTCATGTTCTATGAGCCGATACCAAGTA 1095
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 ThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGlyTyrGlnVal 40
QY 1096 GAAAAAGTGTCTTCATCCAAATATGACTCCAGCAAGCAACAATGACATGACATTGCGCTG 1155
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 GlnLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsnAspIleAlaLeu 60
QY 1156 ATGAAGCTGCAGAGACCTCTGACTTTCAACGACCTAGTGAACACAGTGTCTGCCCAAC 1215
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 MetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsn 80
QY 1216 CCGAGCATGATGCTGCAGCCAGACAGCTCTGCTGGATTTCCGGTGGGGGCCACCGAG 1275
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 ProGlyMetMetLeuGlnProGluGlnLeuCysTrpLysSerGlyTrpGlyAlaThrGlu 100
QY 1276 GAGAAAGGAGACCTCAGAAAGTGTCAACGCTGCCAAGGTGCTTCTCATTTGACACACAG 1335
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 GluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeuIleGluThrGln 120
```

QY 1336 AGATGCAACAGCAGATATGTTCTATGACAACTGATGATCACCAGCCAGTATGTCGCGGC 1395
Db 121 ArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGly 140
QY 1396 TTCTCTGAGGGAAGCTCGATTCTTGCAGGGTGCAGTGGAGGGCCCTCTGGTCACT 1452
Db 141 PheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyProLeuValThr 159

RESULT 9

US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TAGD-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2

Alignment Scores:

Pred No.: 8, 62e-61 Length: 455
Score: 780.50 Matches: 180
Percent Similarity: 55.6% Conservative: 50
Best Local Similarity: 43.5% Mismatches: 158
Query Match: 24.5% Indels: 25
DB: 3 Gaps: 13

US-09-323-597C-1 (1-1738) x US-09-261-416-2 (1-455)

QY 379 ACCCTGGGAGCTTCTCTGCGGAGCTGGCTGGCGGCTGCTACTCTGGAAGTTGATG 438
Db 54 SerLeuGlyIleAlaLeuIleAlaLeuAlaLeuAlaLeuGlyLeuGlyIleHisPhe--- 72
QY 439 GCGCAGCAAGTCTCCAACTCTGGGATAGAGTGGCACTCTCTCAGTACTCTGATCAACCCC 498
Db 73 -----AspCysSerGlyLys---TyrArgCysArgSerSerPheLysCysIleGluLeu 89
QY 499 TCTAACTGGTGTATGGCGGTGTACACTGCCCCGGGGAGGACGAGAAATCGTGTGTT 558
Db 90 IleThrArgCysAspGlyValSerAspCysLysAspGlyGluAspGluTyrArgCysVal 109
QY 559 CGCTCTACGACCAAACTTCTATCTCTCAGGTGTACTCTATCTCAGAGGAAGTCTCTGGCAC 618
Db 110 ArgValGlyGlnAsnAlaValLeuGlnValPheThrAla-----AlaSerTrpLys 127
QY 619 CCTGTGTCAGGACGACTGGAACGAGAACTACGGGGGGGGCGGCTGCGAGGACATGGGC 678
Db 128 ThrMetCysSerAspTrpLysGlyHisTyrAlaAsnValAlaCysAlaGlnLeuGly 147
QY 679 TATAAGATAATTTTACTCTAGCCAGGATAGTGGATGACAGGGATCCACGAGTTT 738
Db 148 PhePro-----SerTyrValSerAspAsnLeuArgValSerSerLeuGluGlyGlnPhe 166
QY 739 ATGAACCTGAACACAAGTGGCGCAATGTC-----GATATCTATATAAAACTGTAC 789
Db 167 ArgGluGluPheValSerIleAsnHisLeuLeuProAspAspLysValThrAlaLeuHis 186
QY 790 CACAGT-----GATGCGCTGTCTTCAAAAGCGAGTGTCTTCTTACGCTGTATA 837
Db 187 HisSerValTyrValArgGluGlyCysAlaSerGlyHisValValThrLeuGlnCysThr 206

QY 838 GCTTGGGGGTCAACTTGAATCAAGCCCGCAGAGCAGGATTTGGCGCGCGAGAGCGCG 897
Db 207 AlaCysGlyHisArgArgGlyTyrSer-----SerArgIleValGlyGlyAsnMetSer 224
QY 898 CTCGGGGGGCTGGCGCTGGCAGGTGAGCTGCAGCTCCAGACCGTCCAGCGTGTGCGGA 957
Db 225 LeuLeuSerGlnTrpProTrpGlnAlaSerLeuGlnPheGlnGlyTyrHisLeuCysGly 244
QY 958 GGTCTCATCATCAACCCCGAGTGGATCGTGACAGCGCCGCTGCTGGGAAAAACCTCTT 1017
Db 245 GlySerValIleThrProLeuTrpIleThrAlaAlaHisCysVal---TyrAspLeu 263
QY 1018 AACATCCATGCGCAATGGAGCGCAATTTGGGGGATTTTGAGACAATCTTTCATGTTCTAT 1077
Db 264 TyrLeuProLysSerTrpThrIleGlnValGlyLeuVal-----SerLeuLeuAspAsn 281
QY 1078 GAGAGC---GGATACCAAGTAGAAAAAGTATTTCTCATCCAAATATGATCTCCAGACC 1134
Db 282 ProAlaProSerHisLeuValGluLysIleValTyrHisSerLysTyrLysProLysArg 301
QY 1135 AAGAACAAATGACATTCGGCTGATGAAGCTGCAGAGCCCTGACTTTCAAGGACCTAGTG 1194
Db 302 LeuGlyAsnAspIleAlaLeuMetLysLeuAlaGlyProLeuThrPheAsnGluMetIle 321
QY 1195 AAACCAAGTGTCTGCCCCAACCCAGGCATGATCTGCAGCCAGACAGCTCTCTCTGGATT 1254
Db 322 GlnProValCysLeuProAsnSerGluGluAsnPheProAspGlyLysValCysTrpThr 341
QY 1255 TCCGGTGGGGGCCACCGAGGAAAGGAGAGACCTCAGAGTGTCTGACGCTGCCAAG 1314
Db 342 SerGlyTrpGlyAlaThrGluAspGlyGlyAspAlaSerProValLeuAsnHisAlaAla 361
QY 1315 GTGCTTCTCTATTCAGACACAG-----AGATGCAACAGCAGATATGCTATGACAACTG 1368
Db 362 ValProLeuIleSerAsnLysAspLeuGlnProGlnGlyArgValArgTrpHisHisLeu 381
QY 1369 ATCACACCGCCCATGATCTGTCGGGCTCTCTCCAGGG---AACGTGATTTCTGCCAG 1425
Db 382 -----ProLeuHisAlaLeuArgGlyLeuProAspGlyTrpArgTrpAsnSerCysGln 399
QY 1426 GGTGACAGTGGAGGGCTCTGGTCACTTCGAAACAACATATCTGGTGGCTGATAGGGAT 1485
Db 400 GlyAspSerGlyProLeuValCysGlnGluArgLeuTrpLysLeuValGlyAla 419
QY 1486 ACAAGCTGGGGTCTCTGCTGTGCCAAAGCTTACAGACCGAGGTGTACGGGAATGTGATG 1545
Db 420 ThrSerPheGlyIleGlyCysAlaAspValAsnLysProGlyValTyrThrArgValThr 439
QY 1546 GTATTACGAGCTGGATTATTCGACAAATGAGGGCGAGAC 1584
Db 440 SerPheLeuAspTrpIleHisGluGlnMetGluArgAsp 452

RESULT 10

US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF SCREENING FOR COLORECTAL CANCER, COMPOSITIONS, A
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.0

; SEQ ID NO 2
 ; LENGTH: 423
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-656-002-2

Alignment Scores:

Pred. No.: 2,99e-52 Length: 423
 Score: 684.00 Matches: 165
 Percent Similarity: 50.22% Conservatives: 64
 Best Local Similarity: 36.18% Mismatches: 167
 Query Match: 21.52% Indels: 60
 DB: 4 Gaps: 16

US-09-323-597C-1 (1-1738) x US-09-656-002-2 (1-423)

QY 292 TCCAAACCCCTGCTGTCAGCCAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG 351
 Db 2 SerAsnPro-----CysAlaAsnProValSerPro---TrpArgProSerGluSerVal 18
 QY 352 ACTAAGAAAGCAGCTGTCATCACCTTGACCCCTGGGACCTTCCTCGTGGAGCTGGCTG 411
 Db 19 GlyIleProLeuIleAlaLeuLeuSerLeuAlaSerIleIleIleValValValLeu 38
 QY 412 GCCCTGCGCTACTCTCGAGTTCATGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 471
 Db 39 IleIysValIleLeuAspLysTyr-----PheLeuCys 50
 QY 472 GACTCCTCAGTACTCATCAACCCCTTAACCTGTTGATGCGGTGTCACACTGCCCC 531
 Db 51 GlyIleProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCysPro 70
 QY 532 GCGCGGAGCAGCAATCGGTGTGCTGCTCTAC-----GGACCA----- 573
 Db 71 LeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAlaVal 90
 QY 574 -----AATCTCATCTTCAGGTGTACTCATCTCAGAGCAAGTCTGTCAC 618
 Db 91 ArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrpPhe 110
 QY 619 CCTGTGTCGAAGCAGTGAAGCAGACTACCGGGGGGGGGCTGCGGACATCGGC 678
 Db 111 SerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysR-glnMetGly 130
 QY 679 TAT-----AAGAAATATTTTACTCTAGCCAA-----GGATA 711
 Db 131 TyrSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeuAspVal 150
 QY 712 GTGATGACACGGATCCACAGCTTTTATGAACATGAACAGTCCCGGCAATGTGAT 771
 Db 151 ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGly----- 167
 QY 772 ATCTATAAAACTGTACCAGTGTGCTGCTTCTTCAAAAGCAGTGGTTCTTTACGC 831
 Db 168 -----ProCysLeuSerGlySerLeuValSerLeuHis 178
 QY 832 TGTATAGCTCGGGGTCAACTGAATCAAGCCCGCAGACAGATGTGGCGGGCAG 891
 Db 179 CysLeuAlaCysGlyLysSerLeu-----LysThrProArgValValGlyGlyGlu 195
 QY 892 AGCGCGTCCCGGGCTGGCCCTGCGAGCTGTCAGCTCCAGAGTCCAGCTG 951
 Db 196 GluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGluHisVal 215
 QY 952 TCGGAGGCTCCATCATCACCCCGAGTGTGATGTCAGCGCCCACTGTGCGTGAAGAAA 1011
 Db 216 CysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLys 235
 QY 1012 CCTCTTAACATCATGCGATGAGCGGCATTTCCGGGGATTTTGAGA---CAATCTTTC 1068
 Db 236 His---ThrAspValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySerPhe 254
 QY 1069 ATGTTCTATGAGCGCGGATACCAAGTAGAAAAAGTGATT-----TCTCATCCAAAT 1119

Db 255 -----ProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnProMet 270
 QY 1120 TATGACTCCAGACCAAGAACATGACATTGGCTGATCAAGCTGCAGAGGCTCTGACT 1179
 Db 271 Tyr-----ProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThr 287
 QY 1180 TTCAAGACCTAGTGAACACAGTGTCTGCCCAACCCAGGCAATGCTGCAGCCAGAA 1239
 Db 288 PheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluIleuThrProAla 307
 QY 1240 CAGCTCTGCTGATTTTCCGGTGGGGGCCACCCAGAGAGAAA---GGGAAGACCTCAGAA 1296
 Db 308 ThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAsp 327
 QY 1297 GTGCTGAACCTGCCAAGGTCTTCTCATTTGACACACAGATGACACAGATGATGTC 1356
 Db 328 IleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAla 347
 QY 1357 TATGACAACTGATCACACAGCCATGATCTGTCGGGCTTCTCGAGGGGAAACGTCGAT 1416
 Db 348 TyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAsp 367
 QY 1417 TCTTGCACGGGTGACAGTGGAGGCTCTGCTCATTCTCGAAGAACAAATATCTGCTGCTG 1476
 Db 368 ThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln---TrpHisVal 386
 QY 1477 ATAGGGGATACAGCTGGGGTCTGCTGTCGCCAAAGCTTACAGACCAGGAGTGTACGG 1536
 Db 387 ValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThr 406
 QY 1537 AATGTGATGGTATTCACGAGCTGATTTATTCGACAAATGAGGGCAGAC 1584
 Db 407 LysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGlu 422

RESULT 11

US-09-851-588-6
 ; Sequence 6, Application US/09851588
 ; Patent No. 6682890

GENERAL INFORMATION:

; APPLICANT: Mack, David
 ; APPLICANT: Gish, Kurt C.
 ; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AI
 ; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
 ; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
 ; CURRENT APPLICATION NUMBER: US/09/851,588
 ; CURRENT FILING DATE: 2001-09-24
 ; PRIOR APPLICATION NUMBER: US 09/642,252
 ; PRIOR FILING DATE: 2000-08-17
 ; PRIOR APPLICATION NUMBER: US 09/656,002
 ; PRIOR FILING DATE: 2000-09-06
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 406

TYPE: PRT

ORGANISM: Homo sapiens

US-09-851-588-6

Alignment Scores:

Pred. No.: 1,35e-51 Length: 406
 Score: 676.50 Matches: 150
 Percent Similarity: 53.91% Conservatives: 57
 Best Local Similarity: 39.06% Mismatches: 128
 Query Match: 21.28% Indels: 49
 DB: 4 Gaps: 13

US-09-323-597C-1 (1-1738) x US-09-851-588-6 (1-406)

QY 508 TGTGATGGCTGTCACTACATCCCGGGCGGAGCAGAGATCGGTGTGCTGCTCTAC 567
 Db 46 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysSerPhe 65

568 -----GGACCA-----AACTTCATCTTCAGGTGTAC 594
Db |||||
66 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 85
595 TCATCTCAGAGAGAGTCTCGCACCTGTGTGCCAAGACGACGTGAAGACGAGAACTACGGG 654
Db |||||
86 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAsnPheThrGluAlaLeuAla 105
655 CGGGCGGCTCGAGGACATGGCTAT-----AAGATAATTTTACTCTAGCCAA-- 705
Db |||||
106 GluThrAlaCysArgGlnMetGlyTySerSerLysProThrPheArgAlaValGluLeu 125
706 -----GGAAATGTGATGACACCGGATCCACACGCTTTATGAACATG 747
Db |||||
126 GlyProAspGlnAspLeuValValGluLeuThrGluAsnSerGlnGluLeuArgMet 145
748 AACCAAGTGGCGCAATGTCGATATCTATAAAACGTACACAGNATGCTGTCT 807
Db |||||
146 ArgAsnSerSerGly-----ProCysLeu 153
808 TCAAAGCAGTGGTTCTTTACGCTGTATAGCTGCGGGGTCAACTTGAACCTCAAGCGC 867
Db |||||
154 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 170
868 CAGACGAGATTCGGGGCGGAGCGCGCTCCCGGGCGCTGCGCTGCAGGTGAGC 927
Db |||||
171 ThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSer 190
928 CTGACAGTCCAGACGTCACGCTGTGCGAGGCTCCATCATCACCCTCGAGTGGATCGTG 987
Db |||||
191 IleGlnTyAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 210
988 ACAGCGCCCTGCTGGTGGAAACCTCTTAACTATCCATGCGCATGACCGGCTTGGC 1047
Db |||||
211 ThrAlaAlaHisCysPheArgLysHis-----ThrAspValPheAsnTrpLysValArgAla 229
1048 GGGATTTTGGAG--CAATCTTTCATGTCTTATGAGCGCGGATACCAAGTAGAGAAAGTG 1104
Db |||||
230 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 245
1105 ATT-----TCTCATCCAAATATGACTCCAGACCAAGACCAATGATGCGGTG 1155
Db |||||
246 IleIleIleGluPheAsnProMetTy-----ProLysAspAsnAspIleAlaLeu 262
1156 ATGAGCTGCAGACCTCTGACTTTCAACGACCTAGTGAACACGAGTGTCTGCCAAC 1215
Db |||||
263 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 282
1216 CCAGCATGATCTGCACGACAGACGCTCTGCTGTGATTTCCGGTGGGGGCCACCGAG 1275
Db |||||
283 PheAspGluLeuLeuThrProAlaThrProLeuTrpIleLeuGlyTrpGlyPheThrLys 302
1276 GAGAAA-----GGAGACCTCAGAGTGTGACCGCTGCAAGGTGCTTCTTCATTGAGACA 1332
Db |||||
303 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 322
1333 CAGAGATGCAACAGCAGATATGCTTATCAACACCTGATCACCACGACCATGATCTGCC 1392
Db |||||
323 ThrArgCysAsnAlaAspAlaTyGlnGlyGluValThrGluLysMetMetCysAla 342
1393 GCTTCTCGAGGGAACTGATCTTCGAGGTGACAGTGGAGGCTCTGCTGCTACT 1452
Db |||||
343 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTy 362
1453 TCGAAGACAAATATCTGTGCTGATAGGGATACAACTGGGGTTCGTGGCTGGCCAAA 1512
Db |||||
363 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyGlyCysGlyGly 381
1513 GCTTACAGACGAGAGTGTACGGAAATGTGATGATTTCAGGACTGGATTTATCGACAA 1572
Db |||||
382 ProSerThrProGlyValTyThrLysValSerAlaTyLeuAsnTrpIleTyAsnVal 401

QY 1573 ATGAGGCGAC 1584
Db 402 TrpLysAlaGlu 405
RESULT 12
US-09-008-271A-6
Sequence 6, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6
Alignment Scores:
Pred. No.: 1,39e-51 Length: 435
Score: 676.50 Matches: 150
Percent Similarity: 53.91% Conservative: 57
Best Local Similarity: 39.06% Mismatches: 128
Query Match: 21.28% Indels: 49
DB: 3 Gaps: 13
US-09-323-597C-1 (1-1738) x US-09-008-271A-6 (1-435)
QY 508 TGTGATGGGTGTACACTGCCCGGGGAGGACGAGAAATCGTGTGTCCCTCTAC 567
Db |||||
75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 94
568 -----GGACCA-----AACTTCATCTTCAGGTGTAC 594
Db |||||
95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114


```
QY 595 TCATCTCAGAGGAGTCTCTGGCCCTGTGTGTGTCACAGACGACTGGAAACGAGAACTACGGG 654
Db 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPhetheGluAlaLeuAla 134
QY 655 CGGCGCGCTCGCAGGACATGGGCTAT-----AAGAATAATTTTACTCTACCCAA--- 705
Db 135 GluThrAlaCysArgGlnMetGlyTySerSerSerLysProThrPheArgAlaValGluIle 154
QY 706 -----GGAATAGTGGATGACAGCGGATCCACCGCTTTATGAAACTG 747
Db 155 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 174
QY 748 AACCAAGTCCGCGCAATGTCGATATCTATAAATACTGTACCACAGTGCCTCTTCT 807
Db 175 ArgAsnSerSerGly-----ProCysLeu 182
QY 808 TCAAAAGCAGTGTCTTTACCTGTATAGCTCGGGGTCAACTTGAACCTCAACCCGC 867
Db 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyGluSerLeu-----Lys 199
QY 868 CAGAGCAGGATTCTGGCGCGCAGAGCGCGCTCCCGGGCGCTGGCGCTGGCAGGTACG 927
Db 200 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 219
QY 928 CTGCACGTCACAGACGTCACAGTGTGGAGGCTCCATCATCATCACCCCGAGTGGTGG 987
Db 220 IleGlnTyAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 239
QY 988 ACAGCGCCGACCTGGTGGAAAACTCTTAACATCCATGCGATGGAGCGCATTTGGG 1047
Db 240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 258
QY 1048 GCGATTTCGAGA---CAATCTTTTCATGTTCTATGAGGCGCGATACCAAGTAGAAAAAGT 1104
Db 259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
QY 1105 ATT-----TCTCATCCAAATTATGACTCCAGACCCAGAACATGACATTCGCGTG 1155
Db 275 IleIleIleGluPheAsnProMetTyr-----ProlAspAsnAspIleAlaLeu 291
QY 1156 ATGACGCTCAGAGGCTCTGACTTTCACAGCCTAGTGAACACCTGTCTGCGCCAC 1215
Db 292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
QY 1216 CCAGCATGATGCTGCAGCCAGAACAGCTCTGCTGATTTCCGGGTGGGGGCCACCGAG 1275
Db 312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 331
QY 1276 GAGAAA---GGGAGACCTCAGAGTGTCTGACCGTCCCAAGTGTCTTCTATTGAGACA 1332
Db 332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
QY 1333 CAGAGATGCAACAGCAGATATGCTATGACAACTCATCACACCCATGATCTGTGCG 1392
Db 352 ThrArgCysAsnAlaaspAspAlaTyrgInGlyGluValThrGluLysMetMetCysAla 371
QY 1393 GCTTCCTCAGGGGAACGTGATCTTCGAGGTGACAGTGGAGGCGCTCTGGTCACT 1452
Db 372 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
QY 1453 TCGAAGAACATATCTGGGGCTGATAGGGGATACAGCTGGGGTCTCGCTGTGCCAAA 1512
Db 392 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrgCysGlyGly 410
QY 1513 GCTTACAGACCCAGGAGTGTACGGGAATGTGATGGTATTCCAGGACTGGATTTATCGACAA 1572
Db 411 ProSerThrProGlyValTyThrLysValSerAlaTyLeuAsnTrpIleTyArgVal 430
QY 1573 ATGAGGGCAGAC 1584
Db 431 TrpLysAlaGlu 434
```

RESULT 13

US-09-851-588-8

; Sequence 8, Application US/09851588

; Patent No. 6682890

; GENERAL INFORMATION:

; APPLICANT: Mack, David

; APPLICANT: Gish, Kurt C.

; APPLICANT: Wilson, Keith E.

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

; FILE REFERENCE: A-68823-1/DJB/JJD/AMS

; CURRENT APPLICATION NUMBER: US/09/851,588

; CURRENT FILING DATE: 2001-09-24

; PRIOR APPLICATION NUMBER: US 09/642,252

; PRIOR FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: US 09/656,002

; PRIOR FILING DATE: 2000-09-06

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-851-588-8

Alignment Scores:

Pred. No.:	1,398-51	Length:	437
Score:	676.50	Matches:	150
Percent Similarity:	53.91%	Conservative:	57
Best Local Similarity:	39.06%	Mismatches:	128
Query Match:	21.28%	Indels:	49
DB:	4	Gaps:	13

US-09-323-597C-1 (1-1738) x US-09-851-588-8 (1-437)

```
QY 508 TGTGATGGGTGTCTACACTGCCCCGGCGGGAGGAGAGATCGGTGTGTCGCTCTAC 567
Db 77 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 96
QY 568 -----GACACA-----AACTTCATCTCTCAGGTGTAC 594
Db 97 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 116
QY 595 TCATCTCAGAGAACTCTCTGCACCTGTGTGCCAAGACGACTGGAAGACGAACTACGGG 654
Db 117 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPhetheGluAlaLeuAla 136
QY 655 CGGCGCGCTCGACGACATGGCTAT-----AAGAATAATTTTACTCTAGCCAA--- 705
Db 137 GluThrAlaCysArgGlnMetGlyTySerSerLysProThrPheArgAlaValGluIle 156
QY 706 -----GGAATAGTGGATGACAGCGGATCCACCGCTTTATGAAACTG 747
Db 157 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 176
QY 748 AACCAAGTCCGCGCAATGTCGATATCTATAAATACTGTACCACAGTGCCTCTTCT 807
Db 177 ArgAsnSerSerGly-----ProCysLeu 184
QY 808 TCAAAAGCAGTGTCTTTTACGCTGTATAGCTCGCGGGTCAACTTGAACCTCAACCCGC 867
Db 195 SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys 201
QY 868 CAGAGCAGGATGTGGCGCGGAGAGCGCGCTCCCGGGCGCTGCGCTCGGAGGTCTAGC 927
Db 202 ThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSer 221
QY 928 CTGCACGTCACAGACGTCACAGTGTGGAGGCTCCATCATCACCCCGAGTGGATCTGTG 987
Db 222 IleGlnTyAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 241
QY 988 ACAGCGCCGACCTGCTGGAAAAAACCCTTTAAACATCCATGCGATGGACCGCATTTGCG 1047
Db 242 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 260
```

```
QY 1048 GGGATTTTGA---CAATCTTTCATGTTCTATGAGCCCGCATACCAAGTAGAAAAAGTG 1104
    |||
    |||
    |||
Db 261 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 276
    |||
    |||
    |||
QY 1105 ATT-----TCTCATCCAAATATGACTCCAGACCAAGAACAAATGATCGCGTG 1155
    |||
    |||
    |||
Db 277 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 293
    |||
    |||
    |||
QY 1156 ATGAAGCTGCGAAGACCTCTGACTTTCACAGCTAGTGAACCAAGTGTGTCTCCCAAC 1215
    |||
    |||
    |||
Db 294 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 313
    |||
    |||
    |||
QY 1216 CCAGGCATGATGCTGCACGACAGACGCTGCTGCTGATTTCCGGTGGGGGGCCACGAG 1275
    |||
    |||
    |||
Db 314 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 333
    |||
    |||
    |||
QY 1276 GAGAAA---GGGAAGACCTCAGAAAGTGTGAAACCTGCCAAGGTGCTTCTCATGAGACA 1332
    |||
    |||
    |||
Db 334 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 353
    |||
    |||
    |||
QY 1333 CAGAGATGCAACAGCAGATATGTTCTATGACAACTGATCACCAGCCATGATCTGGCC 1392
    |||
    |||
    |||
Db 354 ThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAla 373
    |||
    |||
    |||
QY 1393 GGCTTCTCTGCGAGGGAACTGCTGATTTCTGCCAGGTGACAGTGGAGGCGCTCTGCTCACT 1452
    |||
    |||
    |||
Db 374 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 393
    |||
    |||
    |||
QY 1453 TCGAAGAACATATCTGTGCTGATAGGGATACAGCTGGGTTCTGGCTGGCCAAA 1512
    |||
    |||
    |||
Db 394 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTrpGlyCysGlyGly 412
    |||
    |||
    |||
QY 1513 GCTTACAGACAGGAGTGTACGGAACTGTGATGTTACCGGACTGGATTATCGCAA 1572
    |||
    |||
    |||
Db 413 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 432
    |||
    |||
    |||
QY 1573 ATGAGGGCAGAC 1584
    |||
    |||
    |||
Db 433 TrpLysAlaGlu 436
    |||
    |||
    |||
```

RESULT 14

US-08-200-900A-2

; Sequence 2, Application US/08200900A

; Patent No. 5665566

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

; NUMBER OF SEQUENCES: 38

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: MA

; COUNTRY: USA

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/200,900A

; FILING DATE: 23-FEB-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Meinerdt, Maureen C.

; REGISTRATION NUMBER: 31,544

; REFERENCE/DOCKET NUMBER: GI 5201-FWC

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 876-1170 X8574

; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 2:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Alignment Scores:
Pred. No.: 5,01e-50 Length: 798
Score: 660.00 Matches: 150
Percent Similarity: 48.98% Conservative: 67
Best Local Similarity: 33.86% Mismatches: 180
Query Match: 20.76% Indels: 46
DB: 1 Gaps: 12
```

US-09-323-597C-1 (1-1738) x US-08-200-900A-2 (1-798)

```
QY 304 GTCTGCACGACGACCCAAATCCCATCCGACAGTGTGCACCTCAAGACTAAGAAGCA 363
    |||
    |||
    |||
Db 374 ValTyrThrGlyProGlyProValAsnAspValPheSerThrThrAsnArgMetThrVal 393
    |||
    |||
    |||
QY 364 CTGTGCATCACCTTGACCCCTG-----GGGACCTTCTCTCGTGGGA 402
    |||
    |||
    |||
Db 394 LeuPheIleThrAspAsnMetLeuAlaLysGlnGlyPheLysAlaAsnPheThrThrGly 413
    |||
    |||
    |||
QY 403 GCTGCGCTGCGCGCTGCTCTCTGAAAGTTCATCGGCAGCAAGTCTCCAACTCTGGG 462
    |||
    |||
    |||
Db 414 TyrGlyLeuGly-----IleProGluProCysLysGluAspAsn 426
    |||
    |||
    |||
QY 463 ATAGACTGGACCTCTCAGTACTCATCAACCCCTCTAACTGGTGTGATGCGGTGCA 522
    |||
    |||
    |||
Db 427 PheGlnCys---LysAspGlyGluCysIleProLeuValAsnLeuCysAspGlyPhePro 445
    |||
    |||
    |||
QY 523 CACTGCCCCGGCGGAGGACGAGTCTGCTGTGTCTGCTCTACCGA----- 570
    |||
    |||
    |||
Db 446 HisCysLysAspGlySerAspGluAlaHisCysValArgLeuPheAsnGlyThrThrAsp 455
    |||
    |||
    |||
QY 571 -----CCAAACTTTCATCTTCAGTGTACTCATCTCAGAGGAGTCTCTGGCAC 618
    |||
    |||
    |||
Db 466 SerSerGlyLeuValGlnPheArgIleGln-----SerIleTyrHis 479
    |||
    |||
    |||
QY 619 CCTGTGTGCGACAGACTGCGACGAGAACTACGGCGCGCGCGCTCGAGGACATGGCG 678
    |||
    |||
    |||
Db 480 ValAlaCysAlaGluAsnTrpThrThrGlnIleSerAspValCysGlnLeuLeuGly 499
    |||
    |||
    |||
QY 679 TATAAGATAAATTTTACTCTAGCAAGGATAGTCGATCAGACGCGATCCACGAGTTT 738
    |||
    |||
    |||
Db 500 LeuGlyThrGly-----AsnSerSerValProThrPheSerThrThrGlyGlyProTyr 517
    |||
    |||
    |||
QY 739 ATGAAACTGAACAAGTCCCGCATGTCGATATCTATATAAAACTGTACACAGTGTAT 798
    |||
    |||
    |||
Db 518 ValAsnLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrProSerGln 534
    |||
    |||
    |||
QY 799 GCCTGTTCTTCAAAAGCAGTGGTTTCTTTACGCTGT-----ATAGCTCGCGGGTCAAC 852
    |||
    |||
    |||
Db 535 GlnCysLeuGluAspSerLeuIleLeuGlnCysAsnTyrLysSerCysGlyLysLys 554
    |||
    |||
    |||
QY 853 TTGAACCTCAAGCCGCCAGAGC---AGGATTGTGGCGCGGAGAGCCGCTCCCGGGGGCC 909
    |||
    |||
    |||
Db 555 LeuValThrGlnGluValSerProLysIleValGlyGlySerAspSerArgGluGlyAla 574
    |||
    |||
    |||
QY 910 TGGCCCTGCGCAGCTCAGCTGCGACGTCACAGACGTCACGCTGTGCGGAGGCTCCATCATC 969
    |||
    |||
    |||
Db 575 TrpProTrpValValAlaLeuTyrPheAspAspGlnGlnValCysGlyAlaSerLeuVal 594
    |||
    |||
    |||
QY 970 ACCCCGAGTGTGATCGTACAGCCGCCCTCTGCTGTGAAAAAACCCTTTAACTCATCGT 1029
    |||
    |||
    |||
Db 595 SerArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnMetGluProSer 614
    |||
    |||
    |||
QY 1030 CATTGACGCGCATTTCCGGGATTTTGAGACAACTCTTTCATGTTCTATGAGCCGATAC 1089
    |||
    |||
    |||
Db 615 LysTrpLysAlaValLeuGlyLeuHisMetAlaSerAsnLeu-----ThrSerPro 631
    |||
    |||
    |||
```

Qy 1090 CAAGTAGAA-----AAAGTGAATTTCTCATCCAAATATGACTCCAGACC 1134
 Db 632 GlnileGluThrArgLeuileAspGlnileValileAsnProHisTyrAsnLysArgArg 651
 Qy 1135 AAGAACATGATGATGCGCTGATGAAGCTGCAGAGCCTCTGACTTTCAAGCCTAGTG 1194
 Db 652 LysAsnAsnAspIleAlaMeCMeHisLeuGluMetLysValAsnTyrThrAspTyrile 671
 Qy 1195 AAACAGTGTGTCTGCCCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGGATT 1254
 Db 672 GlnProileCysLeuProGluGluAsnGlnValPheProGlyArgIleCysSerile 691
 Qy 1255 TCCGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTCTGCAAGCTGCCAAG 1314
 Db 692 AlaGlyTrpGlyAlaLeuileTyrGlnGlySerThrAlaAspValLeuGlnGluAlaAsp 711
 Qy 1315 GTGCTTCTCATGAGACAGAGATGCAACAGACAGATATGCTATGACCAACCTGATCACA 1374
 Db 712 ValProLeuLeuSerAsnGluLysCysGlnGlnGlnMetProGluTyrAsn---IleThr 730
 Qy 1375 CCAGCCATGATCTGTGGGCTTCTCCAGGGGAAAGCTGATCTTCCAGGCTGACAGT 1434
 Db 731 GluAsnMetValCysAlaGlyTyrGluAlaGlyValAspSerCysGlnGlyAspSer 750
 Qy 1435 GGAGGGCTCTGGTCACTTCCAGAACCAATATCTGGTGGCTGATAGGGGATACAGCTGG 1494
 Db 751 GlyGlyProLeuMetCysGlnGluAsnAsnArgTrpLeuLeuAlaGlyValThrSerPhe 770
 Qy 1495 GGTTCCTGGTGTGCAAGCTTACAGACCAGAGTGTACGGGAATGATGATTCACG 1554
 Db 771 GlyTyrGlnCysAlaLeuProAsnArgProGlyValTyrAlaArgValProArgPheThr 790
 Qy 1555 GACTGGATT 1563
 Db 791 GlnTrpIle 793

RESULT 15
 PCT-US94-00616-2
 ; Sequence 2, Application PC/TUS9400616
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
 ; NUMBER OF SEQUENCES: 33
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (ZPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/00616
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 798 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 PCT-US94-00616-2

Alignment Scores:
 Pred. No.: 5.01e-50 Length: 798
 Score: 660.00 Matches: 150
 Percent Similarity: 48.98% Conservative: 67
 Best Local Similarity: 33.86% Mismatches: 180
 Query Match: 20.76% Indels: 46
 DB: 5 Gaps: 12

US-09-323-597C-1 (1-1738) x PCT-US94-00616-2 (1-798)
 Qy 304 GTCTGCACGGCCCAATCCCATCGGGACAGTGTGCACCTCAAAGACTAAGAGCA 363

Db 374 ValTyrThrGlyProValAsnAspValPheSerThrThrAsnArgMetThrVal 393
 Qy 364 CTGTGTCATCACCTTGACCTG-----GGGACCTTCTCTCGTGCGGA 402
 Db 394 LeuPheIleThrAspAsnMetLeuAlaLysGlnGlyPheLysAlaAsnPheThrThrGly 413
 Qy 403 GCTGCGCTGGCCCTGCTCTGAGAGTTCATCGGCAGCAAGTCTCCAACTCTGGG 462
 Db 414 TyrGlyLeuGly-----IleProGluProCysLysGluAspAsn 426
 Qy 463 ATAGAGTGGACCTCTCAGTACCTCAGTACCTCAACCCCTCTAACTGCTGATGCGGTGCA 522
 Db 427 PheGlnCys---LysAspGlyGluCysIleProLeuValAsnLeuCysAspGlyPhePro 445
 Qy 523 CACTGCCCCGGGGGAGACGAGAAATCGGTGTGTCTCTACCGA----- 570
 Db 446 HisCysLysAspGlySerGluAlaHisCysValArgLeuPheAsnGlyThrThrAsp 465
 Qy 571 -----CCAAACTTCTCAGTCTCAGGTGTACTCTCAGAGGAAGTCTTGCAC 618
 Db 466 SerSerGlyLeuValGlnPheArgIleGln-----SerIleTrpHis 479
 Qy 619 CTGTGTGCAAGACGACTGGAACGAGAACTACGGCGCGCGCTCGAGGACATGGGC 678
 Db 480 ValAlaCysAlaGluAsnTrpThrThrGlnIleSerAspAspValCysGlnLeuLeuGly 499
 Qy 679 TATAAGNAATATTTTACTCTAGCCCAAGGATAGTGTGATGACAGCGGATCCACCAGCTT 738
 Db 500 LeuGlyThrGly-----AsnSerSerValProThrPheSerThrGlyGlyGlyProTyr 517
 Qy 739 ATGAAATCTGAACACAAAGTCGGCAATGTCTGATATCTATAAAACCTGTACACAGTGT 798
 Db 518 ValAsnLeuAsnThrAlaProAsnGlySerLeuIle-----LeuThrProSerGln 534
 Qy 799 GCCTGTCTTCAAAACAGTGTCTTTCAGCTGT-----ATAGCTCGCGGGTCAAC 852
 Db 535 GlnCysLeuGluAspSerLeuIleLeuLeuGlnCysAsnTyrLysSerCysGlyLysLys 554
 Qy 853 TTGAACCTCAAGCGCCAGAGC---AGGATTGTGGCGGAGAGCGCGCTCCCGGGGGCC 909
 Db 555 LeuValThrGlnLeuValSerProLysIleValGlySerAspSerArgGluGlyAla 574
 Qy 910 TGGCCCTGCGAGCTGACGCTGCACAGCTCCAGACAGTCCAGTGTGCGGAGGCTCCATC 969
 Db 575 TrpProTrpValValAlaLeuTyrPheAspAspGlnGlnValCysGlyAlaSerLeuVal 594
 Qy 970 ACCCGCGAGTGGATCTGACAGCGCCCTGCTGTGGAAGAACTCTTAAACATCCATCG 1029
 Db 595 SerArgAspTrpLeuValSerAlaAlaHisCysValTyrGlyArgAsnMetGluProSer 614
 Qy 1030 CATTTGACGGCCATTTGCGGGGATTTTGAGACAATCTTTCATGTCTTATGAGCGCGGATC 1089
 Db 615 LysTrpLysAlaValLeuGlyLeuHisMetAlaSerAsnLeu-----ThrSerPro 631
 Qy 1090 CAAGTAGAA-----AAAGTGAATTTCTCATCCAAATATGACTCCAGACC 1134
 Db 632 GlnileGluThrArgLeuileAspGlnileValileAsnProHisTyrAsnLysArgArg 651
 Qy 1135 AAGAACATGATGATGCGCTGATGAAGCTGCAGAGCCTCTGACTTTTCAAGCCTAGTG 1194
 Db 652 LysAsnAsnAspIleAlaMetHisLeuGluMetLysValAsnTyrThrAspTyrile 671
 Qy 1195 AAACAGTGTGTCTGCCCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGGATT 1254
 Db 672 GlnProileCysLeuProGluGluAsnGlnValPheProGlyArgIleCysSerile 691
 Qy 1255 TCCGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTCTGCAAGCTGCCAAG 1314
 Db 692 AlaGlyTrpGlyAlaLeuileTyrGlnGlySerThrAlaAspValLeuGlnGluAlaAsp 711
 Qy 1315 GTGCTTCTCATGAGACAGAGATGCAACAGACAGATATGCTATGACCAACCTGATCACA 1374

Tue Jul 6 15:26:52 2004

```

Db      712 ValProLeuLeuSerAsnGluLysCysGlnGlnGlnMetProGluTyrAsn---IleThr 730
QY      1375 CCAGCCATGATCTGTGCGGCTTCCTGTCAGGGGAACGTCGATCTTGCAGGTCACAGT 1434
Db      731 GluAsnMetValCysAlaGlyTyrGluAlaGlyValAspSerCysGlnGlyAspSer 750
QY      1435 GGAGGGCCTCTGGTCACCTTGGAAAGCAATATCTGGTCGTCGATAGGGGATACAGCTGG 1494
Db      751 GlyGlyProLeuMetCysGlnGluAsnAsnArgTrpLeuLeuAlaGlyValThrSerPhe 770
QY      1495 GGTTCTGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGTTATTCACG 1554
Db      771 GlyTyrGlnCysAlaLeuProAsnArgProGlyValTyrAlaArgValProArgPheThr 790
QY      1555 GACTGGATT 1563
Db      791 GluTrpIle 793

```

Search completed: July 6, 2004, 15:09:29
 Job time : 42 secs